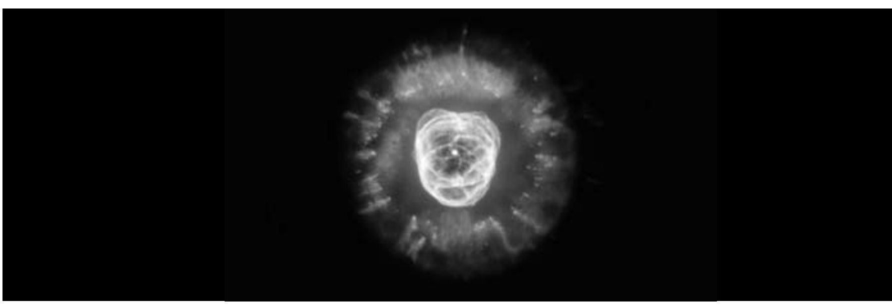
SCIENCE & CERTAINTY

JOHN T.O. KIRK

SCIENCE & CERTAINTY

SCIENCE & CERTAINTY



JOHN T.O. KIRK

CSIRO
PUBLISHING

National Library of Australia Cataloguing-in-Publication entry

Kirk, John T. O. (John Thomas Osmond), 1935.

Science and certainty.
Bibliography.
Includes index.
ISBN 9780643093911 (pbk.).

1. Science – Philosophy. 2. Religion and science.
3. Environmentalism – Philosophy. I. Title.

303.483

Published by
CSIRO PUBLISHING
150 Oxford Street (PO Box 1139)
Collingwood VIC 3066
Australia

Telephone: +61 3 9662 7666
Local call: 1300 788 000 (Australia only)
Fax: +61 3 9662 7555
Email: publishing.sales@csiro.au
Web site: www.publish.csiro.au

Front cover
The Eskimo Nebula (NGC 2392).
Image credit: NASA, Andrew Fruchter and the ERO team.

Set in 10.5/16 Baskerville and Stempel Scheinder
Cover and text design by James Kelly
Typeset by Barry Cooke Publishing Services
Printed in Australia by Ligare

Contents

Preface		vii
Acknowledgements		x
Chapter 1	Origin of the Cosmos and the Solar System	1
Chapter 2	Origin of the Biosphere and *Homo sapiens*	33
Chapter 3	The Nature of Physical Reality	81
Chapter 4	The Physical Basis of Conscious Reality	115
Chapter 5	Environmentalism	143
Chapter 6	Science and Environmental Concern	161
Chapter 7	Gaia	199
Chapter 8	'-isms'	213
References and further reading		243
Index		253

Preface

To some people science is an object of endless fascination and excitement, one of the greatest cultural achievements of the human race. To others, however, science is a source of deep unease. They are fearful of the technologies it has spawned: of their effects on the environment, of the dangers they may sometimes present to human life itself, of their unforeseen and uncontrollable social consequences. To many people, of even greater concern are the possible implications of science for the meaning of human existence, for mankind's view of itself, and for man's significance on this planet and in the universe as a whole.

Although I unhesitatingly admit to being one of those to whom science is a source of enormous interest and satisfaction, nevertheless I acknowledge that the concerns which others have about it are frequently well based, and must be confronted. What I have set out to do in this book is to explore some of these areas of concern: to ask what

in fact does science have to say about some of the most important questions that trouble thinking men and women. What, in short, does science ask us to believe? I will argue that the scientific world-view should be accepted, but that such acceptance in no way commits us to a bleak materialist view of the nature of existence.

Another reason for writing this book was concern about the anti-scientific mood that has developed in some parts of the environmental movement: the tendency to demonize science, and the technologies to which it has given rise, as the source of all the problems which beset the human environment. Believing as I do that our very real environmental problems can only be solved by using science, I regard the anti-science movement as a serious environmental threat in its own right, one which must vigorously be resisted.

The structure of the book is to pose, one after the other, what we might call – the Big Questions. Where do we come from? What is the nature of physical reality? And so on. In each of the first four chapters I take up one of these questions and outline what science, as it exists in this first decade of the 21st century, has to say about it. I make no claim that science can provide complete clear-cut answers to all the 'Big Questions', but it can go a long way towards answering some of them, and can cast light on others. In the following three chapters I enter the problematic area of the relationship between science and environmentalism. After discussing the meaning of 'environmental quality', and outlining the history, and varying philosophies, of the movement, I present what in my view are the major environmental problems confronting humanity, and outline both the science underlying the problems, and the contribution that science can make towards solving them. I follow with what I hope is a judicious, but not unsympathetic, account of James Lovelock's 'Gaia' hypothesis, and its relevance to environmental concerns. In the final chapter I address the thorny problem of the implications of science for some of humanity's more cherished belief systems, and argue that while some should be abandoned, the outlook overall is not as bleak as is commonly assumed.

This is not an exhaustive treatment. Some readers may find that certain topics dear to their heart are absent. This may be because I think they have been thoroughly dealt with by others, or perhaps simply because I did not think they were very important. The choice of other topics for analysis may seem idiosyncratic. For this I make no apology: while the book is intended to address what I believe to be a general need, nevertheless it does in its particular emphases in large part reflect my own range of interests and concerns.

It will be apparent to the reader that the topics addressed in this book cover a wide range of scientific disciplines. My research experience during the half-century or so I have spent as a scientist has been unusually wide, ranging from biochemistry and molecular biology of plants to the physics of solar radiation in aquatic ecosystems, with occasional forays into areas such as plant breeding and grape vine climatology. This does not mean that I am an expert on everything. It does mean that there is no area of science from which I feel excluded, or which I am unwilling to examine. I hope therefore that as the reader accompanies me in this exploration of science and human sensibility, that he or she will feel, regardless of whether they agree with my views, that I am at least not likely to lead them seriously astray on the science of the matter.

John Kirk
Clonakilla Vineyard
Murrumbateman, New South Wales
www.kirkmarineoptics.com

Acknowledgements

I would like to express my thanks to the late Dr Gavin Byrne, and to Dr Michael Kirk, for comments on an earlier version of the manuscript, and to Dr Mark Andrews for helpful comments on Chapter 3. For permission to use figures I wish to thank the NASA/WMAP Science Team (Figs 1.2 and 1.7), NASA (Fig. 1.6), Drs Rolf Schauder and Mark Schneegurt (Fig. 2.2), CALM (Fig. 2.6), Dr Akira Tonomura (Fig. 3.5), and the USDA Natural Resources Conservation Service (Fig. 6.2).

Chapter 1

Origin of the Cosmos and the Solar System

And God said, 'Let there be light', and there was light.

Genesis

Cosmologists are often wrong, but never in doubt.

Lev Landau, Russian physicist (Nobel Prize, 1962)

Of all the questions about the world over which human beings have pondered since they first became capable of rational thought, none are more fundamental than – where do we come from? How did the world come to be the way it is? How did the physical universe which we see displayed around us on any clear night, with its countless billions of stars and galaxies extending to unthinkably remote distances, come into being? How, on one otherwise unremarkable star, did a planet form with equable temperatures, ample liquid water and an atmosphere of benign gases? How, on that planet, did there emerge life forms and ecosystems that in the extraordinary variety of their nature, and the complexity of their interactions, make the merely physical universe seem a thing of comparative simplicity?

Science makes the bold and amazing claim that it can largely answer these questions. It does not claim to explain why there should be a universe

in the first place, or why it should have just this particular set of physical laws and these particular values of the fundamental physical constants. What science says is that given that the physical universe does exist, then by invoking the physical laws whose operation has been revealed by careful observation and experiment, and using the measured values of various fundamental properties of matter and energy, it is possible to infer the character of the very much simpler state from which the universe began in the very distant past, and explain how it developed from that very simple condition to the enormously rich and complex state in which we find it.

In this and the following chapter I present an outline of what seems to me to be the best scientific understanding of how the physical and biological worlds as we know them have come to be. My account will be simplified and brief, and since there are excellent books on these topics available I will only summarize the evidence upon which this understanding is based. In the last chapter we shall consider the implications of this for humanity.

ORIGIN OF THE UNIVERSE

Cosmologists – physical scientists whose field of interest is the cosmos as a whole – believe that the universe came into existence somewhere between 10 and 20 thousand million years ago. Recent estimates seem to be zeroing in on something close to 13.7 billion years. This conclusion is based on the fact, known since the work of US astronomer Edwin Hubble in the 1920s, that the universe is expanding. The galaxies are steadily moving apart from each other; they do so not because they have been hurled apart by some force in a pre-existing space, but because the very fabric of space itself is growing. From careful measurements of the rate of expansion – analysis of the spectrum of the light from distant galaxies tells us how fast they are moving – it is possible to calculate backwards to the time when all the matter in the universe started off at the same point. This works out to be

about 14 billion years before the present. This initial event, the appearance of an unimaginably concentrated pinpoint of matter and energy ~13.7 billion years ago which was to give rise to our universe, is colloquially referred to as the 'Big Bang'. Although the term 'Big Bang' is now thoroughly entrenched, it is inappropriate in one respect in that it gives the impression that the galaxies are moving apart because they were blown away from each other by some primal explosion. It thus obscures the underlying reality that space itself is expanding and simply carrying the galaxies along with it.

The universe is presently filled with radiation photons. These are of very low energy; they correspond to photons which would be emitted by a very cold body only 2.7 K above absolute zero (Fig. 1.1). These photons constitute the so-called background radiation of the universe, and on the basis of its omnipresence and uniformity the universe can be said to have an average temperature of 2.7 K. If we now imagine going back in time, i.e. reversing

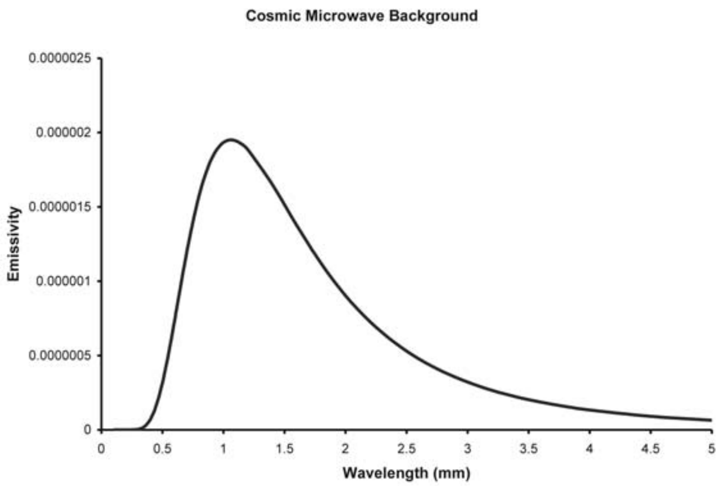


Fig. 1.1 Spectral energy distribution of cosmic microwave background. The energy (as emissivity in watts m^{-2} mm^{-1}) of this left-over radiation from the Big Bang is here plotted as a function of wavelength in mm. It corresponds to that of a black body at 2.73 K.

the expansion, then as space gets smaller the wavelength of the radiation also becomes smaller, and necessarily (see Chapter 3) its energy rises. In addition, as space shrinks so the number of photons per unit volume increases, and as a consequence of the increase in both the energy and the concentration of the photons the temperature of the universe rises. For most of our journey backwards the universe remains on average a very cold place, but as we get near to the beginning and we have compressed our radiation into a much smaller space, the temperature of the universe begins to rise steeply. For example, when we are about 400 000 years out from time zero the temperature is already 3000–4000 K. When we get as close as 100 seconds to the Beginning the temperature is up to 1 billion (10^9) K, and if we go all the way back to 1 microsecond (10^{-6} s) the temperature is in the region of 1–10 thousand billion (10^{12}–10^{13}) K.

All these extraordinary figures can be calculated quite straightforwardly from classical physics on the basis of what happens when you compress a known amount of radiation into a smaller and smaller volume. High temperatures can be equated to high energies (in electron-volts) of any particle existing at those temperatures; experimental studies by particle physicists during the last 100 years have provided a great deal of information on how subatomic particles behave at very high energies. The most powerful particle accelerators constructed to date have given us information on particle behaviour at the energies that would have existed in the universe at t~1 microsecond, although not at the even higher temperatures that would have prevailed earlier.

Thus from about $t = 10^{-6}$ s onwards a coherent and physically plausible, although hypothetical, account of the growth and development of the physical cosmos can be given. Just before we embark on this journey, in the forward direction this time, let us very briefly consider what can be said about the universe before $t = 10^{-6}$ s. In fact nothing can be said with any confidence, but speculation abounds. Some speculations start off with an empty universe and consider whether, with the help of some rather bizarre physics to push matters along, the whole thing might have started off as a

quantum fluctuation in the vacuum.[14,21] There is no reason why such highly speculative proposals should not be explored by theorists. There is, however, an unfortunate tendency to refer to hypotheses that start with the vacuum, as explaining the creation of the universe *ex nihilo*, i.e. from nothing. This is a misuse of language. In reality there is no such thing as empty space. Quantum electrodynamics, invoking Heisenberg's uncertainty principle (Chapter 3) as quoted by physicist Paul Davies,[8] tells us that:

> *What might appear to be empty space is … a seething ferment of virtual particles. A vacuum is not inert and featureless, but alive with throbbing energy and vitality.*

Or again, to quote the cosmologists John Barrow and Frank Tipler in their book, *The Anthropic Cosmological Principle*:[1]

> *It is, of course, somewhat inappropriate to call the origin of a bubble universe in a fluctuation of the vacuum 'creation* ex nihilo*', for the quantum mechanical vacuum is not truly 'nothing'; rather the vacuum state has a rich structure which resides in a previously existing substratum of space-time, either Minkowski or de Sitter space-time. Clearly a true 'creation* ex nihilo*' would be the spontaneous generation of everything – space-time, the quantum mechanical vacuum, matter – at some time in the past.*

Thus the quantum vacuum is anything but empty. It is 'a ferment of quantum activity, teeming with virtual particles and full of complex interactions'.[8] These particles are coming and going and interacting in multidimensional space. If a quantum vacuum existed before the Big Bang, it was certainly not 'nothing': it was already a universe. Cosmological hypotheses of this type therefore do not answer the metaphysical question of why the universe should exist at all.

Current theories of how the universe might have proceeded from an imagined pre-existing quantum vacuum to the Big Bang, are highly speculative. The most popular one is the inflation hypothesis put forward by

Alan Guth[14] and others. This envisages that the universe began as a tiny patch of space in a state of quasi-infinite temperature and density which, after about 10^{-35} seconds, began to expand at an enormous and accelerating rate achieving an expansion of $\sim 10^{26}$ in $\sim 10^{-33}$ seconds. The inflation was driven by a hypothetical *inflaton* field (so called by analogy with other field names ending in 'on', e.g. photon field and gluon field), the fundamental nature of which we presently know nothing about. As the universe underwent this initial enormous expansion it inevitably cooled. At the end of inflation, a process called *reheating* occurred, in which the inflaton particles (with every field there is an associated particle) decayed into the radiation that started the 'standard' hot Big Bang. The attraction of 'inflation' scenarios is that they account for certain otherwise puzzling properties of the observable universe, such as the extreme uniformity of the background radiation – this does not vary by more than one part in 10^5 in any direction. The very rapid expansion would smooth out the distribution of energy. We shall not pursue the inflation hypothesis further, but proceed to the comparatively well-attested Big Bang itself, whose description we will take up at $t = 10^{-6}$ s, when the universe had cooled to the point at which the laws of physics we presently know can be assumed to apply. We shall largely follow the classic description given by Stephen Weinberg in *The First Three Minutes*[22] and the article 'The origin of the universe' by Victor Weisskopf,[23] modified in the light of more recent findings as described in the books by John Gribbin,[13] Armand Delsemme[9] and Simon Singh,[20] supplemented with information from the *Encyclopedia of Astronomy and Astrophysics.*[11]

FROM THE BIG BANG TO THE FIRST ATOMS

The universe, when one microsecond old, was entirely filled with radiant energy in the form of gamma-ray photons. There was very little else. The density of the energy, i.e. its concentration per unit volume of space, was unthinkably enormous. The individual photons of the radiation field were

themselves so energetic that the energy in one photon was greater than the available mass-energy (in terms of the well-known $E = mc^2$ relationship) in certain particle/antiparticle pairs (see the box below for a summary of current understanding of the essential subatomic constituents of matter). Spontaneous conversion of photons to electron/positron, quark/antiquark and neutrino/antineutrino pairs, and to another kind of particle, the gluons, would have taken place. Recombination of quarks with antiquarks, and electrons with positrons, to yield once again radiant energy in the form of gamma-ray photons would also occur, and so interconversion between particles and radiation was going on simultaneously in both directions.

Meanwhile the universe – which, it is thought, was at the time quite small (although no-one seems able to actually put a figure on it) – was rapidly expanding. This at first diminished the rate of matter–antimatter collisions more than it diminished the rate of conversion of gamma-ray photons to particles, so although photons were still vastly in the majority the proportion of the total mass-energy which existed as particles steadily grew.

With expansion the total energy density, and thus the effective temperature, of the universe was decreasing. Some of the particles which were being created directly from radiant energy could combine in various ways to form larger particles. At the very high temperature ($\sim 10^{12}$–10^{13}K) prevailing at 10^{-6} s or earlier, these combinations were unstable but by the time the universe was ~ one-thousandth to one-hundredth of a second old (temperature $\sim 10^{11}$–10^{12} K), some of the larger particles were surviving. Together with the gluons which provide the binding force, quarks or antiquarks were combining in threes to give protons, neutrons, antiprotons and antineutrons, and in twos to give mesons, but such particles were still vastly outnumbered by the photons, electrons, positrons, neutrinos and antineutrinos which still dominated the universe. Once formed, the protons and neutrons underwent continual rapid interconversion by the processes:

$$\text{proton} + \text{antineutrino} \Leftrightarrow \text{neutron} + \text{positron}$$

$$\text{neutron} + \text{neutrino} \Leftrightarrow \text{proton} + \text{electron}$$

Initially these processes went equally rapidly in either direction but with further cooling of the universe the more-energy-demanding conversion of the proton to the slightly heavier neutron became less favoured and the composition of the nucleon mix gradually shifted towards proton dominance.

Neutrinos in the present-day universe interact hardly at all with other matter – they just pass through it as though it is not there. They did interact, as indicated above, in the very early universe, only because of the enormous density that existed at the time. As the early universe expanded its density of course decreased, and some time between 0.1 and 1 seconds from the beginning, interaction between neutrinos/antineutrinos and all the other particles virtually ceased. As it is sometimes expressed, the neutrinos and antineutrinos 'decoupled' from the rest of the universe. The original neutrinos/antineutrinos are all still out there, but are not playing any significant role that we are aware of.

A further consequence of expansion was that the energy of the photons decreased. The stretching of space stretched out the waves, and thus increased the wavelength of the radiation. The energy of photons is, as we shall see in Chapter 3, inversely proportional to their wavelength, so as the average wavelength of the radiation photons increased due to the expansion of space, their energy decreased. A consequence was that the rate of conversion of radiation to particles decreased. The annihilation of matter and antimatter particles, which happened whenever they collided and were converted back to gamma-ray photons continued, and by about $t = 1$ second, annihilation of particles was beginning to occur more rapidly than their re-creation from radiation.

Since exactly equal numbers of matter and antimatter particles are produced when gamma-ray photons transform into particles, the question arises – how is there is any matter left in the universe? Why has it not simply recombined with the antimatter, leaving a universe filled only with radiation? The answer is not known, but may well be a consequence of the fact that when certain unstable particles decay they produce a very slight excess

of matter over antimatter; and it is this proportionately tiny excess which supposedly has given rise to the material world we see around us. For example, it has been shown that when neutral K mesons decay, the process is not quite symmetric and leads to an excess of 1 part in 500 of matter over antimatter in the decay products.

Once the universe had expanded, and consequently cooled enough for protons and neutrons to survive, it became possible for these to start combining to form the atomic nuclei of some of the elements that exist in the world today. At first, because of the high temperature, such combinations would have had only a fleeting existence, but over the approximate period ~1/100 s to 100 s, during which time the temperature fell from $\sim 10^{11}$ to $\sim 10^{9}$ K (1 billion degrees), protons – which were present in excess (88 protons to 12 neutrons, by 100 s) – combined with all the neutrons. Initially, mainly nuclei of deuterium (the heavy isotope of hydrogen) consisting of one proton and one neutron, known as *deuterons*, were formed. By reactions involving additional protons or neutrons virtually all the deuterons were then converted to nuclei of Helium-4 (the common isotope of helium), consisting of 2 protons and 2 neutrons. A very small proportion of the deuterons survived, together with a comparable number of Helium-3 nuclei (2 protons, 1 neutron) and an even smaller number of nuclei of Lithium-7 (4 protons and 3 neutrons, equivalent to a fusion of a Helium-3 and a Helium-4 nucleus). Within 3–4 minutes of the Big Bang, it seems that the universe had settled down to a stable nuclear composition of about 76% hydrogen (a single proton being a hydrogen nucleus) and 24% helium, with just traces of deuterium (0.002%), Helium-3 (0.001%) and Lithium-7 (0.00000001%). This composition was to persist for about the next 200 million years.

In addition to this very limited range (compared to the present) of atomic nuclei, the universe at this stage (~4 minutes) still contained vast numbers of photons (~3 billion per nucleon), neutrinos and antineutrinos which had ceased to interact with anything else, and free electrons exactly equivalent in number to the free or combined protons. Over the next ~370 000

years the universe continued its expansion and associated cooling, but for most of this time was hot enough to remain as a *plasma*, a mixture of free nuclei and free electrons. Any fleeting combinations between the two were almost immediately destroyed. Free electrons cause scattering of photons and so the universe, although glowing white-hot and containing no solid particles whatever, was effectively opaque. Towards the end of this period, however, the temperature had fallen from its initial ~1 billion degrees to a few thousand degrees, and the neutral atoms formed by combination between the negative electrons and the positive nuclei began increasingly to survive. By t = ~370 000 years the temperature had fallen to 3000–4000 K, the free nuclei had combined with electrons and the plasma had been replaced with an electrically neutral gas consisting almost entirely of hydrogen and helium. This crucial transition in the history of the universe is referred to as *recombination*. Hydrogen atoms were combining with each other to form molecular hydrogen (H_2). With the disappearance of the free electrons of the plasma, scattering of the photons – now consisting mainly of visible and near-infrared radiation, with wavelengths in the region of ~0.7 μm (0.007 mm) – ceased and the universe became transparent. The radiation photons remaining from the Big Bang then became, like the neutrinos before them, decoupled from the rest of the universe, i.e. they ceased to interact with matter. Unlike the neutrinos, however, they can still be detected all around us in space, but by now cosmic expansion has stretched them out so much that we have to look in the microwave region of the spectrum at wavelengths of about 1 mm (Fig. 1.1).

FORMATION OF THE GALAXIES AND THE FIRST STARS

For the next 100 million years or so nothing dramatic happened. The universe continued to expand and to cool: by t = ~10^8 years it still consisted of an enormous mass of hydrogen and helium gas which was becoming very cold, with a temperature of ~30 K or less. A hypothetical human

observer present at the onset of transparency at t = ~370 000 years would have seen a uniform red glow; by t = ~100 million years the observer would have seen nothing. The universe now was utterly dark: no radiation visible to the human eye was emitted anywhere. This is sometimes referred to as the Cosmic Dark Age. If our imaginary observer had concluded that this particular universe was going nowhere, that it was destined simply to go on forever as a more and more rarefied cloud of hydrogen and helium with temperature asymptotically approaching absolute zero (0° K) they would have been wrong, because great changes were already in train.

The mass of hydrogen/helium gas left after the onset of transparency was of almost perfectly uniform density. The extent of this departure from uniformity is revealed in the minute fluctuations across the sky in the intensity of the cosmic microwave background (CMB). These fluctuations – no greater than 1 part in 100 000 – have been mapped in great detail by the Wilkinson Microwave Anisotropy Probe aboard a satellite 1.5 million km from the Earth.[18] (Fig. 1.2). The tiny variations in brightness of the CMB correspond to variations in the density of the primordial plasma just before it cooled into a neutral gas. The more compressed regions of the plasma had a higher temperature, and so glowed more brightly. The fluctuations in density are attributed to sound-waves in the plasma which, like sound-waves in air, created alternately compressed and rarefied regions.[15] When recombination suddenly occurred the patterns were frozen into the cosmic microwave background.

Upon the very small inhomogeneities in the distribution of mass in space, the force of gravity began very slowly but inexorably to operate. Any region of space with a greater than average density would act as a centre of gravitational attraction for the gaseous matter around it, and so nearby molecules and atoms would tend slowly to move towards it. In this way gravitation progressively amplified any pre-existing inhomogeneities, and so the near-uniform mass of gas initially present eventually separated out into individual giant clouds surrounded by almost empty space. These cold clouds of hydrogen and helium were destined to become the galaxies.

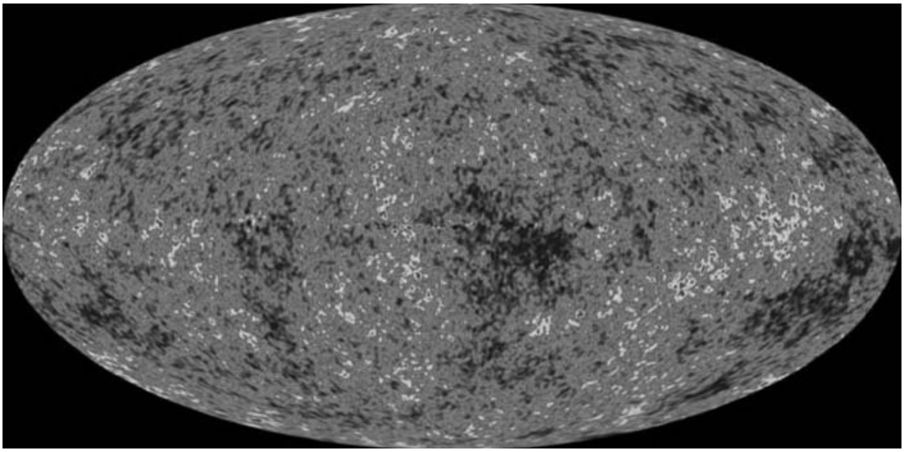


Fig. 1.2 The Microwave Sky (courtesy NASA/WMAP Science Team). The detailed, all-sky picture of the infant universe from three years of WMAP data. The image reveals 13.7 billion year old temperature fluctuations (shown as density differences) that correspond to the seeds that grew to become the galaxies.

Within the individual gas clouds, which were probably approximately spherical in shape, there would have been smaller inhomogeneities of mass distribution upon which gravitation would operate. These developed into globules of gas much smaller than the gas cloud as a whole, but large enough so that by their own inherent gravitational attraction they began to contract in upon themselves. As the mass became more concentrated, the contraction accelerated. The gas molecules, falling faster and faster, acquired kinetic energy; when they arrived at the centre and began to collide with each other this energy was manifested as an increased temperature. Thus the universe, although still cooling overall, began to develop marked inhomogeneities of temperature – local hot-spots. Our imaginary observer, materializing at this time in one of these gas clouds, would have seen points of light appearing here and there in the otherwise black sky.

The temperature at the centre of the contracting globule of gas depended on the total mass of the globule: the greater the mass the higher the central temperature. Provided the mass of gas was at least 8% of the mass of the Sun, by the time gravitational collapse was complete the

temperature at the centre would be not less than 10 million K. This is so hot that when the hydrogen nuclei (protons), by now stripped of their electrons, collided with each other they began to fuse to form helium nuclei. In the overall process there is a slight reduction in mass, and therefore an accompanying release of an equivalent amount of energy:

$$4 \text{ protons} \rightarrow 1 \text{ Helium-4} + 2 \text{ positrons} + \text{energy}$$

Astrophysicists refer to this, somewhat inappropriately, as the 'burning' of hydrogen to give helium. For simplicity we shall do the same.

The collapsed globule of gas now had a self-sustaining energy source. When this began to happen within giant gas clouds, or protogalaxies, around the universe, the first stars were born. These – the first generation of stars, referred to by astronomers as Population III stars – began to appear perhaps as early as ~200 million years after the Big Bang. Unlike later stars they consisted essentially only of hydrogen and helium, and are thought to have been very large, with 100 or more solar masses. Compared to our Sun, with its lifetime of billions of years, they were very short-lived, burning out in just a few million years. Figure 1.3 shows the structure of a typical medium-range star about the same size as our Sun. Although the temperature at the centre of a star is always at least 10 million K, the

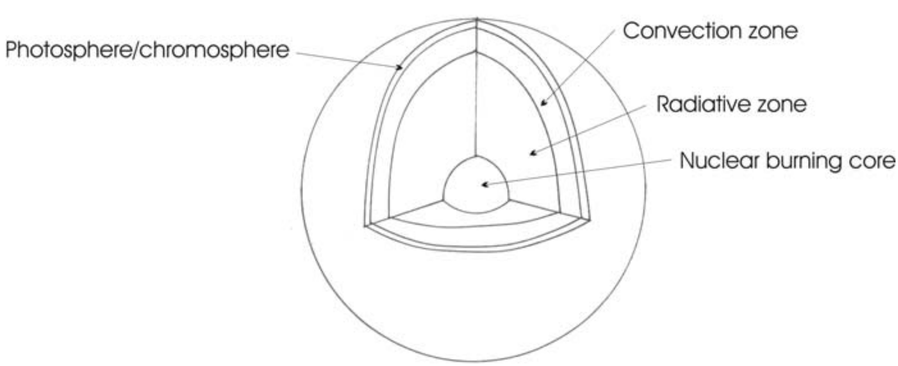


Fig. 1.3 Internal structure of a star such as the Sun (after Crannell, 1989).[11]

temperature at the surface is much less – typically some thousands of degrees, leading to emission of radiant energy predominantly in the visible region of the spectrum. The imaginary observer of our galaxy at this time would have seen something not dissimilar to what we see today – pinpoints of intense white light displayed on the black backdrop of space. Despite the apparent similarity to the present day, the very earliest stars still consisted of essentially nothing but hydrogen and helium. There was a lot of cosmic chemistry to be carried out before the stage was set for the appearance of planets and life.

The great cloud of gas which filled the universe during the Cosmic Dark Age was essentially neutral, i.e. it consisted of intact atoms of hydrogen and helium. It was transparent to visible and near-ultraviolet light, but not to ultraviolet radiation of short wavelengths around 1216 Å (121.6 nm), which hydrogen atoms absorb. The first stars, however, as well as being much bigger than our Sun, were much hotter and emitted mainly in the ultraviolet (UV) region of the spectrum. When a hydrogen atom absorbs one of these highly energetic UV photons, its electron is raised to a higher energy level and frequently flies off altogether, leaving ionized hydrogen, i.e. a bare positively charged proton, behind. Ionized hydrogen no longer absorbs UV. A growing bubble of ionized gas would have formed around each of these first stars.[16] As more stars formed, the bubbles of ionized gas merged until eventually, after about 900 million years, re-ionization of the gas cloud was complete.

FORMATION OF THE CHEMICAL ELEMENTS BY STELLAR NUCLEOSYNTHESIS

To understand how the new chemical elements required for our existence were brought into being we shall follow what happens during the development path (Fig. 1.4) of a large star, after fusion of hydrogen to helium begins. The important nuclear processes in stars all take place within a

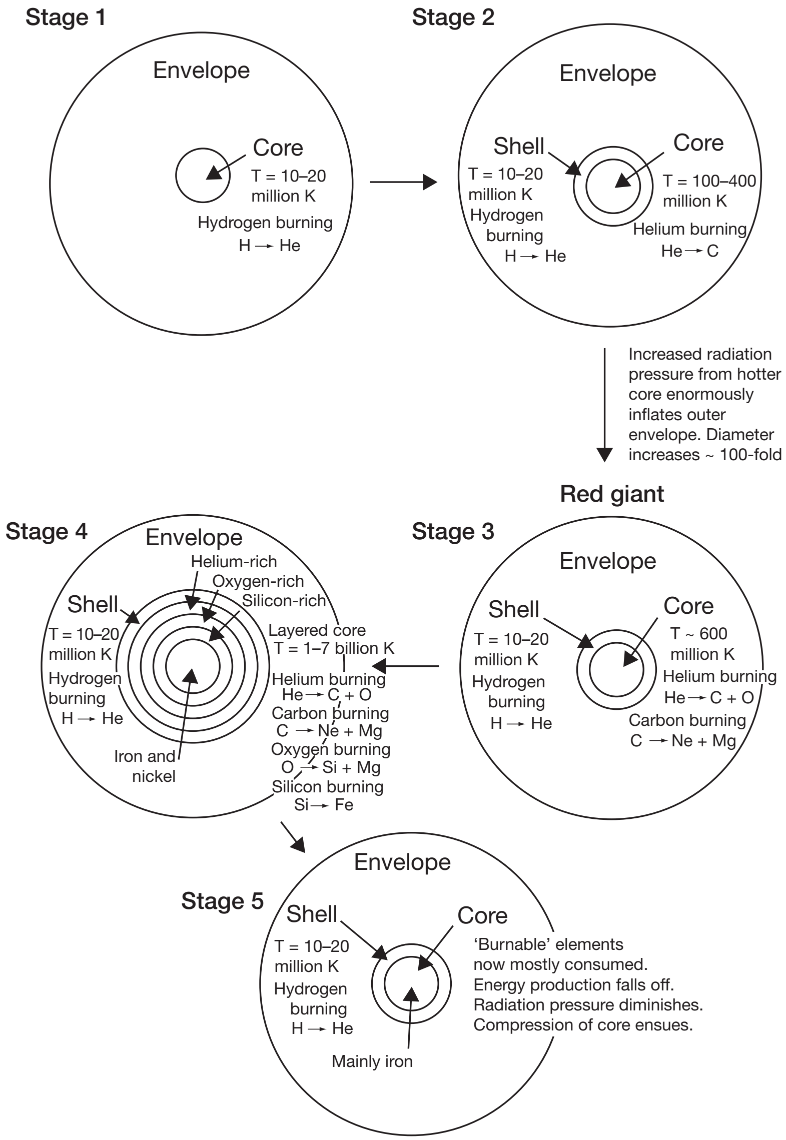


Fig. 1.4 Nucleosynthesis during the life and death of a large star. (a) Sequential stages in stellar evolution. From the hydrogen-burning young star to the fuel-depleted burnt-out star about to collapse.

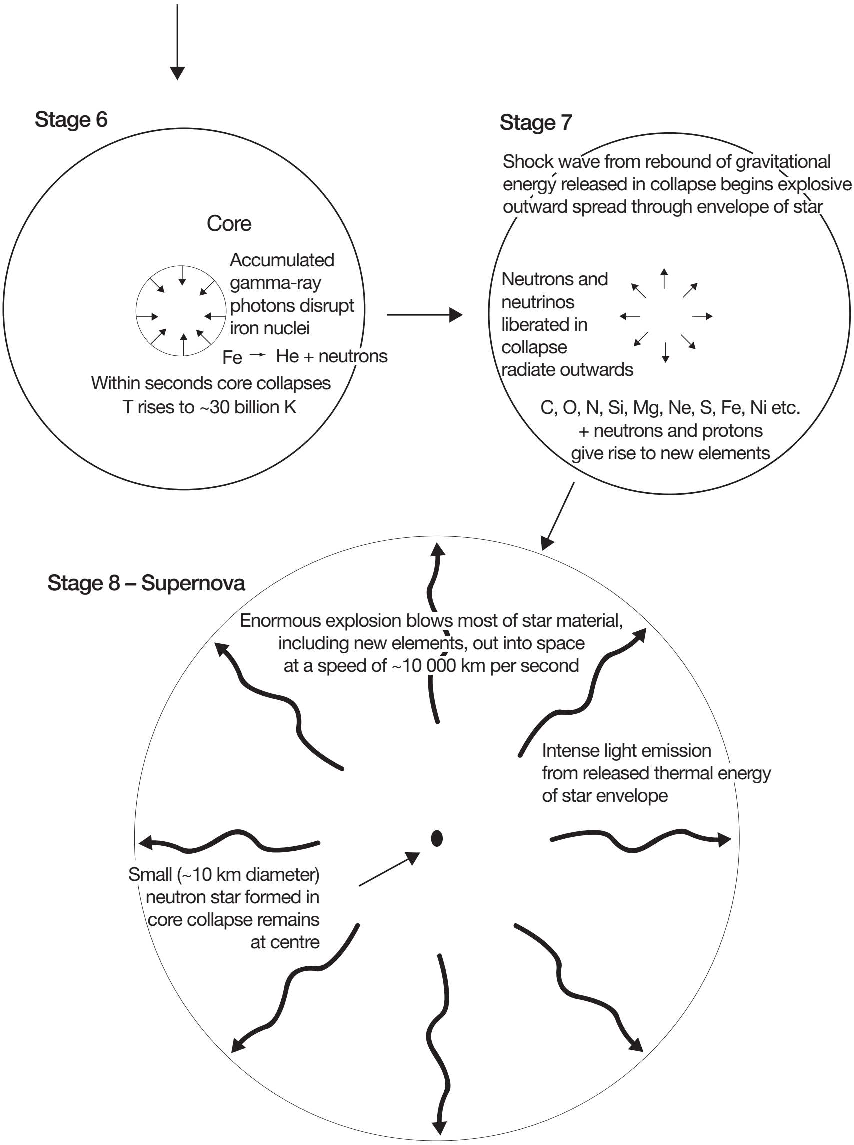


Fig. 1.4 (b) Gravitational collapse of burnt-out star is followed by a supernova explosion in which most of the star material is blown out into space, with the accompanying synthesis of many new elements by combination of the liberated neutrons and protons with the existing nuclei.

region at the centre known as the *core*, which may have a diameter only one-tenth that of the whole star, and a thin concentric *shell* around it. These are surrounded by a much larger 'convective envelope' of low density. To begin with, 'burning' of hydrogen to helium proceeds in the core. Eventually all the hydrogen in the core is used up. The radiation pressure initially falls and the surrounding mass of the star once again begins to collapse towards the centre. As it does so the temperature of the core rises: when it reaches 100 million K, a new energy-liberating reaction becomes possible. This is the combination first of two helium nuclei to give a beryllium nucleus, and the combination of this with another helium nucleus to give a carbon nucleus plus two gamma-ray photons:

$$^{4}He + {}^{4}He \rightarrow {}^{8}Be$$
$$^{8}Be + {}^{4}He \rightarrow {}^{12}C + 2\gamma$$

Once carbon is formed, the carbon nuclei themselves begin to fuse to form oxygen:

$$^{12}C + {}^{4}He \rightarrow {}^{16}O + \gamma$$

This reaction is slow, which is just as well since otherwise all the carbon would soon be consumed (see Chapter 8) and there would be none left to make people out of! Nevertheless, it becomes significant at higher temperatures. This process of 'helium burning' involves an overall diminution of mass and therefore liberation of energy. Meanwhile, in the shell around the core the temperature has reached 10 million degrees, so the burning of hydrogen to helium, previously confined to the core, continues in the thin layer of shell.

As the helium in the core is consumed further contraction occurs, due to the surrounding pressure, and the temperature progressively rises. Other nuclear fusion reactions, creating and also consuming the nuclei of new heavier elements, come into play, proceeding up the periodic table of elements as far as iron, but no further. Iron is the most stable nucleus and there is no more stable nucleus to which it could be converted, by bombardment with protons or helium nuclei, with a resultant liberation of energy.

As the material of the core ends up increasingly as iron, so the fuel for energy production becomes depleted and the radiation pressure holding the great mass of the star at bay decreases. Once again contraction begins but now there are no remaining energy-generating new nuclear reactions which can be switched on to build up the radiation pressure again. With increasing compression the gamma-ray photons liberated in earlier nuclear reactions become so energetic that they begin to react with the iron nuclei, breaking them up into helium nuclei and neutrons. This process soaks up energy so that the temperature of the core falls, triggering a very rapid – on a time-scale of seconds – gravitational collapse of the core under the weight of the surrounding mass of star.

The core at this stage contains iron and helium nuclei and free electrons which were stripped from the hydrogen and helium atoms when the star first began to heat up. The gravitational collapse compresses the electrons so much that they penetrate into the nuclei, where they react with the protons converting them into neutrons. As electrons disappear, the part of the internal pressure which is due to free electrons diminishes and gravitational collapse is further accelerated. Soon most of the matter in the core is in the form of free neutrons, and the core has become a neutron star (even a black hole if the original star was sufficiently massive), a mere 10 km or so in diameter.

Part of the gravitational energy released in the inward collapse develops into a shockwave which, on its rebound from the centre, manifests as an enormous outwards explosion of the surrounding material of the star, thrown off at great velocity into space. There is also intense emission of radiation, some in the visible region of the spectrum. This is the phenomenon of a *supernova*, which for a few days can be much brighter than the whole of the galaxy in which it occurs. For our purpose, however, the main significance is that there is also an intense outward flux of some of the neutrons liberated at an earlier stage. As they travel they collide with nuclei of those elements already formed (carbon, oxygen, nitrogen, silicon etc.) and, together with the protons (hydrogen nuclei) which are already freely

available, bring about synthesis of the remaining elements of the periodic table, i.e. those beyond iron, which as we saw above could not be formed by the nuclear fusion processes in the original star.

The heavy element formation which takes place during a supernova is referred to as *r-process* nucleosynthesis,[6] the *r* standing for *rapid*. It requires a high neutron flux such as occurs briefly during a supernova. Elements heavier than iron are also synthesized in stars by another neutron capture process referred to as *s-process nucleosynthesis*, the *s* standing for *slow*. This takes place within long-lived stars of low or intermediate mass towards the end of their lives: the resulting elements are ejected into interstellar space when these stars finally collapse.

Although observed supernovae have been very rare in human history – only seven have been seen in our galaxy in the last 2000 years – they are in fact quite common galactic events. It has been estimated that in a fairly large spiral galaxy such as ours they occur at an average interval of about 20 years: we fail to see most of them, however, because of our location in the dusty obscuring disk of the Milky Way.[3] As a result of the nucleosynthetic processes both preceding and occurring during these supernovae the galaxy, although still consisting mainly of hydrogen (70%) and helium (28%), eventually came to contain about 2% of other chemical elements, comprising varying amounts of all the (stable) elements in the periodic table. Carbon, nitrogen and oxygen are the major non-H/He elements, with iron, magnesium, silicon, sulphur and neon making up most of the rest. As a result of the violence of the supernova explosions, all these elements became well mixed into the galaxy and so were incorporated into second, third and later generation stars. Thus eventually all the chemical ingredients for the formation not merely of new stars, but of actual planetary systems, were created and distributed. The stage was now set for the formation of our own solar system.

Stardust to Planets: Creation of the Solar System

Within the stars themselves the new elements did not form molecules. At the very high prevailing temperatures they existed as positive ions: bare nuclei in a super-hot plasma of protons (hydrogen nuclei), helium nuclei and free electrons. The nuclei ejected into interstellar space and not incorporated into new stars were able, once they had cooled down, to combine with electrons to give neutral atoms, and these started reacting with each other to form molecules. As well as diatomic hydrogen (H_2) itself, a wide range of simple molecules was formed including water (H_2O), molecular nitrogen (N_2), carbon monoxide (CO), formaldehyde (HCHO), hydrogen cyanide (HCN), carbon dioxide (CO_2), ammonia (NH_3), methane (CH_4), sulphur dioxide (SO_2) and many others. All these can be detected by radio astronomy in our galaxy today. In addition, metal atoms by stepwise combination with oxygen and silicon – two of the more common elements that they would encounter – formed silicates, the building materials of rocks.

At the very low temperatures prevailing in space the kinetic energy of the molecules was so low that when they collided they would sometimes remain together, and in this way very small grains of *interstellar dust* began to form. There are two main categories of dust grains. One consists of iron and magnesium silicates surrounded by a layer of frozen water, with small amounts of carbon dioxide and other organic compounds. The other is composed of carbon (mainly as graphite), reduced carbon compounds (polyaromatic hydrocarbons, metallic carbides) and iron.

Each supernova would give rise to an individual cloud of neutral atoms, molecules and dust grains; as individual clouds were brought together by chance as a result of their movement within the galaxy, gravitational forces led them to tend to remain together. In this way giant interstellar clouds built up. The Great Nebula in the constellation of Orion, and (for southern hemisphere readers) the Coal Sack on the lower left-hand side of the Southern Cross, are examples. The enormous mass within such clouds

means that there is a strong tendency for gravitational contraction to occur, but this is opposed by the turbulent motion of the gas. Mass density varies throughout the cloud, in places it will be high enough for gravitation to prevail and irreversible contraction to begin (Fig. 1.5). The localized 'blob' of gas and dust that is beginning to contract is not internally motionless. Different parts of the mass have a variety of velocities in

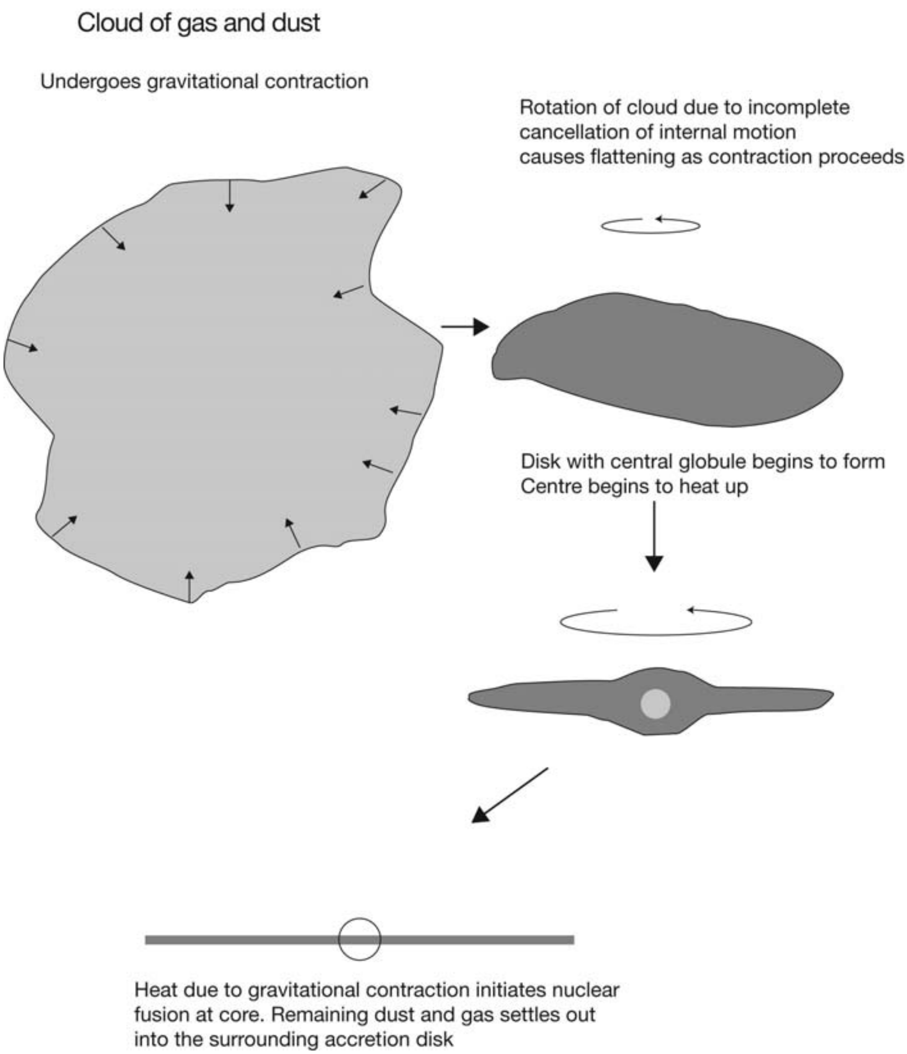


Fig. 1.5 Formation of the Sun and accretion disk from a cloud of gas and dust.

various directions, due to turbulent motion. Except in the unlikely event that these all exactly cancel out, this internal turbulent kinetic energy will, as contraction proceeds, manifest as a net rotational energy or angular momentum, and the blob begins to rotate. As it rotates it begins to flatten out to the shape of a disk with a central globule. With continued contraction, and accretion of matter from the inner part of the disk, the central globule heats up, just as in the formation of the first generation of stars discussed earlier, and eventually turns into a star sustained by nuclear fusion at its core. In this way our own Sun came into existence, but at this stage surrounded only by a disk of gas and dust.

Under the influence of its own gravitation the disk became progressively flatter, leading to increased concentration of the dust grains in the central plane. This in turn led to an increased rate of particle collision, and the consequent formation of aggregates of larger and larger size. Pebble-sized bodies soon became common. During the course of about 10 000 years these aggregates grew into irregular objects ranging in size from 1 to 10 km in diameter;[9] at this stage they are referred to as *planetesimals*. When planetesimals came in contact as they drifted within the disk they tended to adhere together, thus getting larger and larger. Eventually such aggregates became sufficiently large to have significant gravitational attraction of their own. Smaller bodies were drawn in towards them and growth became faster still. Over 100 000–1 million years this led to the formation of *planetary embryos* with masses comparable to the Moon or Mars (0.01–0.1 Earth masses).[5] Further aggregation of these embryos – a few dozen, in the case of the Earth – led to the formation of the planets of our solar system.

The chemical composition of the dust grains in the disk from which the planets were formed varied in accordance with distance from the Sun, because the temperature of the disk increased markedly with increasing proximity to the Sun. At the radius where the Earth was to form, the temperature is thought to have been as high as 1100 K, so that all water and other volatile matter was lost and the residual dust grains consisted of iron and metallic silicates.[9] Thus the early Earth, consisting of iron and silicates,

did not have a chemical composition favourable for the development of life. Furthermore, because of the great heat generated by the numerous gravitationally-induced collisions with other large objects in the final stages of its formation, our planet was actually molten. The silicates were less dense and floated to the top, while the iron concentrated at the centre. This is why the Earth has a rocky silicate crust and a (still molten) iron core.

From the birth of the Sun, surrounded by its disk of gas and dust, to the formation of the Earth together with its satellite the Moon (thought to result from collision of our planet with another massive body) may have taken as little as 30–50 million years. Over the next ~0.5 billion years the Earth was severely bombarded by comets. These were planetesimals formed in the cold outer regions of the solar accretion disk, where the giant planets Jupiter, Saturn, Uranus and Neptune appeared. Because of the lower temperature of their formation, these planetesimals were rich in water, carbon compounds and other volatile constituents of the primeval dust grains. They can be regarded as left over from the outer regions of the accretion disk, after formation of the giant planets was complete. Planetesimals which were not captured by the growing giant planets had their orbits perturbed by the massive gravitational objects to such an extent that they were thrown out of their original orbits in various directions. Many collided with the Earth and its satellite. The scars of this cometary bombardment are readily visible on the surface of the Moon, parts of which are covered with impact craters of all sizes (Fig. 1.6). The large smooth areas (*maria* or 'seas') of the Moon correspond to regions where the heat of cometary impact was sufficient to melt the surface crust, and thus remove all evidence of previous collisions. The Earth does not show such cratering, possibly because the surface melting caused by cometary bombardment was sufficiently extensive to obliterate the craters – there was greater impact velocity due to Earth's greater gravitational pull. Later geological processes of water and air erosion, and tectonic movement of the Earth's crust, have also contributed to the removal of the evidence. With time the rate of cometary bombardment greatly decreased, falling in

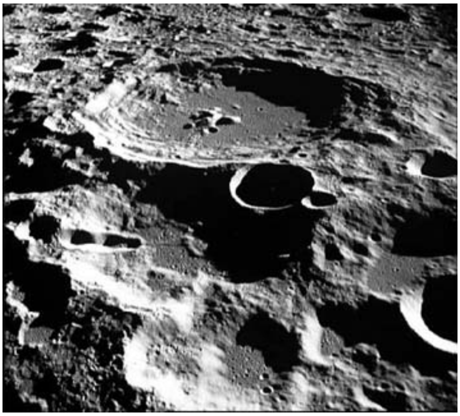


Fig. 1.6 Craters on the Moon. This is an oblique view of the Moon's far side, photographed from the Apollo 11 spacecraft on 16 July 1969. The large crater in the centre of the picture is International Astronomical Union crater no. 308, which has a diameter of about 80 km (50 miles). The rugged terrain is typical of the far side of the Moon.
Source: NASA image; caption based on the original NASA caption released with the image.

the second half-billion years to only ~1% of that in the preceding half-billion years. The cometary authority, Armand Delsemme,[9] estimates that after the first half-billion years of our planet's existence the rate dropped quickly to a few comets per century and then to a few per millennium.

After about half to 1 billion years virtually all the stray planetesimals from the giant planet region had been swept up by various members of the solar system, or had otherwise gone their way. Major cometary impacts with the Earth became rare, the planet's crust from now on remained solid, and temperature had fallen to the point at which water could remain liquid. The next stage in the development of the Earth began. The period of intense cometary bombardment, although it precluded the development of life while it lasted, played a crucial role in preparing our planet for the ultimate appearance of living things. The comets brought water, carbon, nitrogen, sulphur and other necessities of life to the otherwise very inhospitable ball of iron and silicates which was the early Earth.

Our Universe – A Progress Report

Before we leave cosmology and begin our account of how living systems arose on this planet, we will take one last look at the universe itself to see what has become of it since the moment of creation. When, after ~370 000 years, it had cooled down to the point at which nuclei and electrons could stably combine to form atoms, its elemental composition was about 75% hydrogen and 25% helium, with traces of deuterium and lithium. After 13.7 billion years the universe still consists mainly of hydrogen and helium in a 3:1 proportion but now, as the result of repeated cycles of nucleosynthesis within stars followed by ejection into and mixing with the interstellar medium, it contains all the elements in the periodic table, all the way up to thorium and uranium. These other elements supply only about 2% by weight of the chemical composition of the cosmos.

Nearly all the weight of any atom is contributed by the protons and neutrons in the nucleus – these subatomic particles, together with some other rarely encountered particles (e.g. the antineutron), are lumped together as baryons. Although we know the overall chemical composition of the universe very well, unfortunately this does not mean that we know what the universe is made of – there is 5–6 times as much matter which is non-baryonic in nature[2,12]. The existence of this cosmic constituent, referred to as cold dark matter, which has so far not been directly observed, is inferred from a variety of astronomical observations. The first known example came from the observation by astronomer Fritz Zwicky in 1933 that many of the galaxies in the Coma cluster of galaxies were moving so fast that they should have exceeded the escape velocity expected on the basis of the amount of visible matter present. Zwicky concluded that there must also be large amounts of invisible mass holding the cluster together. The stability of our own galactic disk can only be explained on the hypothesis that it is embedded in a more massive spherical mass of unseen material. Observed gravitational lensing of light (the bending of light rays from distant stars by the gravitational pull of a galaxy as they pass close to it) is

invariably much greater than that expected on the basis of the visible matter in the galaxies. The nature of this cold dark matter remains a mystery. Various new kinds of particles have been postulated, but so far there is no evidence for the existence of any of them.

Our ignorance of the universe's composition is not limited to the unknown nature of dark matter. One of the conclusions derived from the detailed measurements of the cosmic microwave background by WMAP and earlier instruments was that the geometry of the universe is very close to being spatially 'flat', i.e. the normal rules of Euclidean geometry apply to the universe as a whole and that light, where it is not affected by gravity, travels in straight lines. General relativity theory shows that this can only be true if the mass density of the universe has a certain specific value. Combining the best estimates of the amounts of visible matter and dark matter provided only about one-third of the requisite density, so about two-thirds of the mass in our universe was still not accounted for.

In 1998 two groups of astronomers – one led by Brian Schmidt of the Mount Stromlo Observatory in Canberra, Australia, and the other by Saul Perlmutter at the Lawrence Berkeley National Laboratory, US – independently reached the extraordinary conclusion that the expansion of the universe, far from slowing down as had commonly been assumed, was in fact accelerating. Both groups were studying supernovae in very distant galaxies. Discrepancies between distances calculated from the redshift of the light (due to expansion of the universe) and those calculated from the observed brightness of the supernovae could only be explained by the hypothesis that the rate of expansion is increasing with time. It is thought that there are two opposing forces which determine the behaviour of the universe over time: gravitational attraction which tends to slow expansion, and some omnipresent unidentified energy field which opposes gravity and promotes expansion. During the universe's initial phase of existence, perhaps as much as half its lifetime, when the galaxies were closer together and the matter density was high, it is thought that gravity was stronger and was slowing the expansion. However, as expansion continued the matter

density progressively diminished and eventually reached the point at which the unknown energy field became stronger. Expansion then began to accelerate.

The unknown energy field is referred to as *dark energy*. It is believed to be smoothly distributed, and of nearly constant value. One candidate is *vacuum energy*, also called the *cosmological constant*, an absolutely constant minimum energy of empty space itself.[4] Energy and mass are equivalent, and the current view is that dark energy supplies the ~two-thirds of missing matter in the universe. The best recent estimate is that the universe consists of 73% dark energy, 23% dark matter and only 4% visible (baryonic) matter.

One question which is often asked about the universe is – how big is it? We know how big the 'visible universe' is. It is a sphere, centred on the Earth, whose radius is the distance light photons have travelled since the Big Bang 13.7 billion years ago. This radius is about 45 billion light-years: although the photons have been travelling for only 13.7 billion years, expansion of the universe in this time has greatly elongated their route.[15] But of course we have no reason to suppose that the physical universe stops just at the periphery of our vision. It seems entirely likely that there are other galaxies, stars and planets out there which we will never see. On the other hand, we have equally no reason to suppose that the universe goes on forever. Despite the occasional loose talk by some cosmologists of an 'infinite' universe or 'infinite' space, actual infinities are purely mathematical entities which exist within set theory but cannot possibly exist within the real world.[7] Quite apart from the philosophical inconceivability of an actually infinite universe, there is a particular feature of the cosmic microwave background, as revealed by WMAP, which can be interpreted as direct evidence of a finite universe.[17,10] The sky map in Figure 1.2 represents the distribution of microwave intensity at the maximum angular resolution of the WMAP, which is about 0.3°. The same data set can also be mapped at any selected level of angular resolution – 1°, 10°, 30° etc., and the average size of the fluctuations in intensity across the sky can be

calculated for each such mapping. When these average fluctuations are plotted against the angular interval, we obtain the *angular power spectrum* (Fig. 1.7), the precise characteristics of which can then be compared with the predictions of various cosmological theories.[19] Although most of the curve fits standard cosmological models, there is a loss of power at large angular scales (>60°) which was not anticipated. Jean-Pierre Luminet and coworkers[17] interpret this as evidence that the very largest sound-waves in the primordial plasma are missing – perhaps because space itself is not big enough to support them. The density parameter, Ω_0 (the ratio of the actual density of the universe to what is known as the *critical density*) calculated from WMAP and other data has the value 1.02 ± 0.02. Taken at its face value, this indicates that the universe is finite and closed.[10]

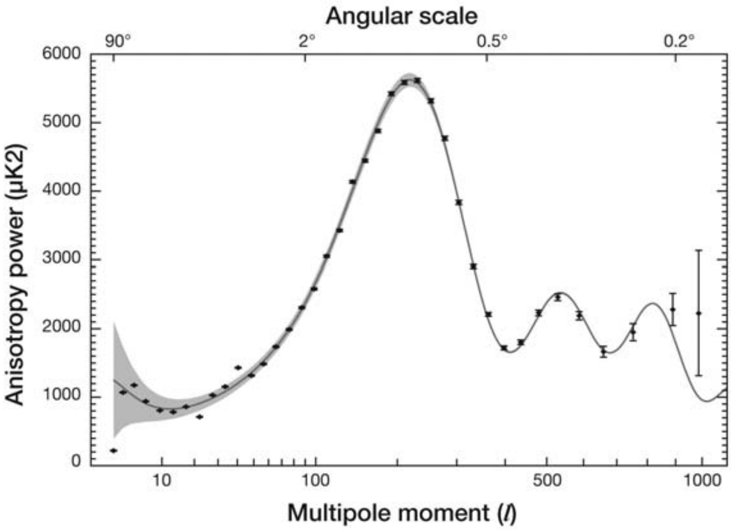


Fig. 1.7 Angular power spectrum of the temperature anisotropy of the cosmic microwave background, calculated from the first three years of measurements by the WMAP satellite.
Source: Courtesy NASA/WMAP Science Team.

SUBATOMIC CONSTITUENTS OF MATTER

A large number of different subatomic particles have been discovered by using very high energy particle accelerators. The vast majority of these, however, are very short-lived, with lifetimes in the region of 10^{-6} to 10^{-23} seconds. Athough the study of these ephemeral objects has been of great importance in revealing the principles upon which all particle structure is based, they are not themselves constituents of normal matter and need not concern us here.

The atoms in the Earth and other planets, and the free nuclei in the incandescent hearts of the Sun and other stars, are all composed of a variety of combinations of a very small number of fundamental building-blocks. The simplest atom, hydrogen, consists of a single proton with a single electron in orbit around it. The electron has a negative charge: the proton has an equal and opposite positive charge, and a mass which is 1836 times that of the electron. The second simplest atom is that of the heavy isotope of hydrogen, deuterium. It has a nucleus consisting of one proton plus one neutron, and this again has a single electron in orbit around it. The neutron is slightly (~0.014%) heavier than the proton and has no electric charge. The nuclei of all the other chemical elements that exist are composed of protons and neutrons in varying numbers and proportions, and in every case the atom of the element has orbiting electrons equivalent in number to the protons in the nucleus.

Protons and neutrons, which are generically referred to as *nucleons*, are composed of subunits known as *quarks* (first hypothesized by Murray Gell-Mann and George Zweig independently in 1963). There are two kinds of quarks in these nucleons, which are called 'up' and 'down' quarks – simply to distinguish them, with no geometrical or functional implications. Taking the electron and the proton as each

having a charge of unity (negative and positive, respectively), the up quark has a charge of $+\frac{2}{3}$ and the down quark a charge of $-\frac{1}{3}$. Each nucleon contains 3 quarks: 2 up and 1 down in the proton, giving rise to a net charge of +1; 1 up and 2 down in the neutron, giving rise to a net charge of zero. The electron, by contrast to the nucleons, appears to have no substructure, no parts out of which it is built. Quarks can also combine in pairs, an up with a down, to give particles known as *mesons*. These are highly unstable and short-lived.

Within the nucleons the quarks are bound together by what is referred to as the *strong nuclear force*. In the modern theory of quantum electrodynamics, which developed out of quantum mechanics, all forces between the elementary subatomic building-blocks of matter are mediated by the continual rapid exchange of 'messenger particles'. In the case of the strong force, the messenger particles which continually exchange between the quarks to hold them together, and which in a similar manner hold the protons and neutrons together in the nucleus, are known as *gluons*. They themselves have no mass and no charge.

As well as matter, the universe consists of radiation. This is electromagnetic energy propagating through space at 3×10^8 m s^{-1}. Depending on the means we use to observe it, it manifests either as wave energy or as particles (Chapter 3). In its particulate aspect, we refer to the particles as *photons*. Photons have no mass. We can see radiation with wavelengths between 0.4 and 0.7 micrometres; this is referred to as the *visible waveband*. At longer wavelengths (lower-energy photons) we have infrared and microwave radiation. At shorter wavelengths we have, at increasing photon energy, UV radiation, X-rays and gamma-rays.

For each of the fundamental particles – electron, proton, neutron, up quark, down quark – out of which normal matter is built,

there exists an 'antiparticle' which is similar in properties, equal with respect to mass but opposite in charge (where there is a charge), and so constituted (simply a fact for which we have no explanation) that when a particle collides with its equivalent antiparticle the two are annihilated, their combined mass-energy being released as pure radiant energy. Conversely, a high energy radiation photon can spontaneously convert into matter: a single gamma-ray photon, for example, can change into an electron accompanied by a positively-charged antielectron, known as the *positron*. Although matter built entirely out of antiparticles could exist in principle, it is now believed that it does not exist: the universe is entirely constituted of normal matter, not antimatter, and antiparticles are seen only when they are created in particle accelerators, in cosmic ray (high energy particles from space) interactions with the atmosphere or in the radioactive decay of unstable nuclei.

There is one further particle which we need to know about. Although, strictly speaking, it is not a constituent of atoms, nuclei or nucleons, it nevertheless appears when unstable atoms or nucleons decay and is common but virtually invisible throughout the universe. This is the *neutrino*, together with its antiparticle the *antineutrino*. When the nuclei of certain radioactive elements such as carbon-14 (^{14}C) undergo *beta decay*, one of the neutrons in the nucleus spontaneously changes into a proton, with the simultaneous emission of an electron (originally called a 'beta particle') and an antineutrino, a nitrogen nucleus being formed in the process:

$$^{14}C \rightarrow {}^{14}N + e^- + \nu$$

When another, much more unstable, isotope of carbon, ^{11}C, decays a proton changes into a neutron, and a positron and neutrino are emitted:

$$^{11}C \rightarrow {}^{11}B + e^+ + \nu$$

It is important to realize that the antineutrino and the neutrino did not escape from confinement within the neutron and proton: they, together with the electron or positron, came into existence when the nucleon underwent its spontaneous change. Neutrinos and antineutrinos were originally thought to have no mass and to travel at the speed of light, but recent experiments indicate that they have a very tiny mass and travel at a little less than the speed of light.

Chapter 2

Origin of the Biosphere and *Homo sapiens*

What is man that thou art mindful of him?

Psalms 8:4

Someone must have been telling lies about the dinosaurs, for without having done anything wrong an asteroid came out of the sky and killed them one fine morning.

Attributed to Franz Kafka

That science, if allowed to assume certain physical laws and constants, can plausibly explain how the physical universe came to be the way it is, is not generally seen as an affront to human sensibility. But science does not leave it at that. It also claims not only to understand in principle, albeit incompletely, how living things work, but also in large part to understand how the whole living world in all its beauty and enormous complexity came to be. In particular, biological science offers an explanation, entirely in terms of natural processes, of how humankind itself arose within the animal world. It cannot be denied that for many people these claims are problematic, even threatening. This is partly because they offer what might seem to be a soulless and mechanical account of the origin of our immensely rich and beautiful world of living things and complex ecosystems. More particularly it is because the scientific view of the biological

world is often used, especially by the inexplicably enthusiastic propagandists for triumphalist materialism, to justify the downgrading of humankind to just one more species, of no more value or significance than any other. In this chapter we shall see just what science says about how the living world, including human beings, came to be. Whether this scientific worldview in fact has the bleak metaphysical implications so often claimed for it is quite another matter, to which we shall return later in the book.

The ~1 billion-year-old Earth, at the point reached at the end of Chapter 1, was now physically and chemically complete. Its gravitational attraction was sufficiently strong to prevent gas molecules escaping into space, so that it maintained a gaseous atmosphere. At 150 million km from our own moderately-sized star, the Sun, the Earth received enough radiation to keep most of the water on the planet liquid, without boiling it off into steam. It had a rich endowment of all the chemical elements in the periodic table, forged in the furnaces of countless galactic supernovae in the preceding ~8 billion years that the universe had existed. The physics and the chemistry had been done. The Earth was at this time completely sterile, but the stage was set for biology to begin.

First, a word on nomenclature. We shall be discussing events at various enormously distant times in the past, and our units of time will typically be billions or millions of years ago. I will use the common abbreviation 'bya' to indicate 'billions of years ago', and 'mya' to indicate 'millions of years ago'.Thus '3.5 bya', for example, means 3.5 billion years ago. The history of our planet since its formation, as revealed by the geological record, shows major changes in the kinds of rock being laid down, or in the living organisms present at various intervals. To make this history more digestible, geologists and palaeontologists have organized the overall timescale into a series of divisions and subdivisions corresponding to the transitions observed. The names of the most relevant (to us) divisions, and the periods of time to which they correspond, are given in Table 2.1.

Table 2.1: The geological time scale of the Earth. The transitions from one division to another generally correspond to major geological or palaeontological events such as mass extinctions revealed by changes in the fossil record. The actual dates of the transitions are inevitably somewhat uncertain. The ones given here are those proposed by the International Commission on Stratigraphy.

Geological time since our planet was formed is divided into four *eons*
Hadean eon: 4.5–3.8 bya
Archean eon: 3.8–2.5 bya
Proterozoic eon: 2.5 bya–542 mya
Phanerozoic eon: 542 mya to the present
The eons are subdivided into eras, e.g. the Proterozoic eon
Paleoproterozoic era: 2.5–1.6 bya
Mesoproterozoic era: 1.6–1.0 bya
Neoproterozoic era: 1.0 bya–542 mya
The Phanerozoic eon consists of the
Paleozoic era: 542–251 mya
Mesozoic era: 251–65.5 mya
Cenozoic era: 65.5 mya to the present
Eras are further subdivided into periods; our concern is the last period of the Neoproterozoic era
Ediacaran period: ~630–542 mya
and various periods within the Phanerozoic eon
Cambrian period: 542–488 mya
Ordovician period: 488–444 mya
Silurian period: 444–416 mya
Devonian period: 416–359 mya
Carboniferous period: 359–299 mya
Permian period: 299–251 mya
Triassic period: 251–200 mya
Jurassic period: 200–145 mya
Cretaceous period: 145–65.5 mya
Tertiary period: 65.5–1.81 mya
Quaternary period: 1.81 mya to the present
The Tertiary and Quaternary periods are subdivided into *epochs*
Tertiary period
Paleocene epoch: 65.5–55.8 mya
Eocene epoch: 55.8–33.9 mya
Oligocene epoch: 33.9–23.0 mya
Miocene epoch: 23.0–5.3 mya
Pliocene epoch: 5.3–1.8 mya
Quaternary period
Pleistocene epoch: 1.8 mya–11 400 years ago
Holocene epoch: 11 400 years ago to the present

THE FIRST ORGANIC MATTER – EARTH'S COMETARY DOWRY

The post-cometary Earth, 3.5–4 bya, had a very dense hot atmosphere, many times the present atmospheric pressure, consisting mainly of carbon dioxide, with some nitrogen, ammonia and hydrogen, and a hot ocean. In addition to its inorganic constituents (CO_2, carbonates, silica, silicates, sulphate, ammonia, nitrate, phosphate, chloride, bromide, various metal ions etc.), the ocean water contained a wide range of small organic molecules. Where did these all come from?

Chemical analysis of meteorites and cometary dust, and spectroscopic analysis of radiation from interstellar clouds, show that simple compounds like methane, formaldehyde, hydrogen cyanide and methanol are formed in space on the grains of interstellar dust, and so are molecular building-blocks of living matter, such as amino acids (subunits of proteins), aliphatic carboxylic acids (related to fats), purines and pyrimidines (components of nucleic acids) and sugars (used in energy metabolism).[9,11,34]

As the icy planetesimals bombarded Earth during the cometary reign of terror, they brought, in addition to water and carbon dioxide, a significant payload of organic molecules (Fig. 2.1). Armand Delsemme[11] estimates that during the ~1 billion years of cometary bombardment as much as 10^{18} tonnes of these 'prebiotic' molecules were deposited on the planet. Since this corresponds to an average deposition of ~200 g per sq cm of the Earth's surface per million years, if the estimate is at least of the right order of magnitude it means that the Earth began its post-cometary career already with a very substantial endowment of bio-relevant organic materials.

In addition to coming from space in the form of comets and cometary dust (which is still falling into our atmosphere today) organic molecules could have been created on Earth, without the involvement of living organisms, by electrical discharges (lightning) within an atmosphere of appropriate composition. In a famous experiment carried out half a century ago, Stanley Miller and Harold Urey subjected a mixture of methane, hydrogen, water vapour and ammonia to a repeated series of electrical discharges in a sealed flask over a period of days. The condensed water was found to

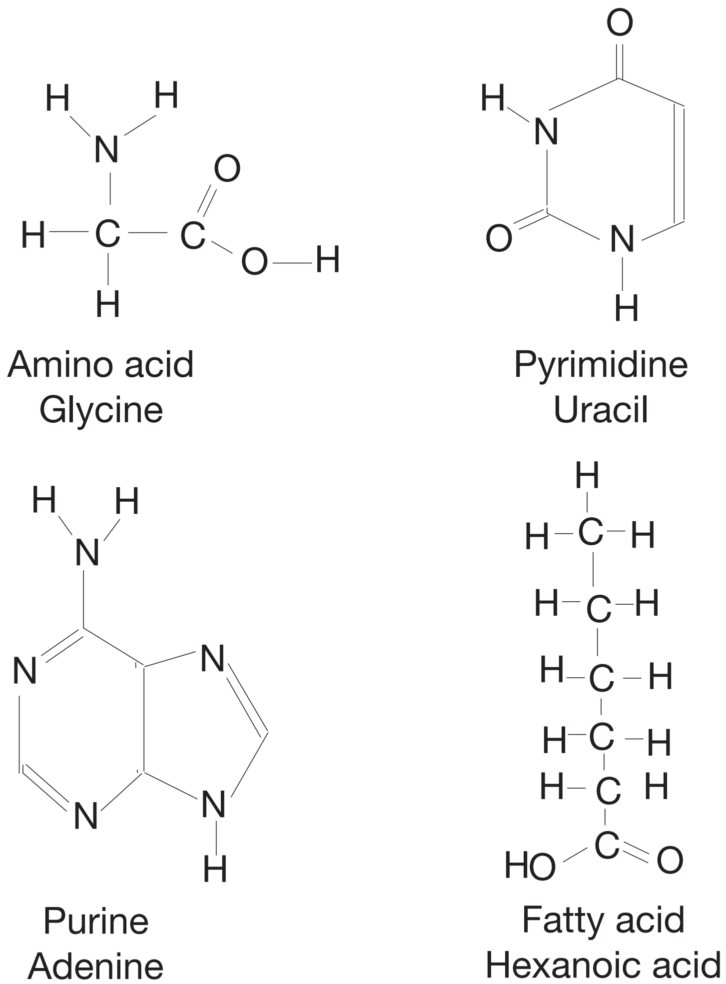


Fig. 2.1 Some of the simple organic molecules which are the building-blocks of life. Proteins consist of chains of amino acids. Nucleic acids consist of chains of nucleotides, each of which contains a purine or pyrimidine base such as adenine or uracil. The key components of the lipid bilayer, the structural principle of all biological membranes, are fatty acids. Organic compounds such as these are formed in space on interstellar dust and were brought to Earth by comets.

contain a complex mixture of organic compounds including some amino acids. A possible alternative, or additional non-biological mechanism for creating organic molecules, suggested by Günter Wächtershäuser[38] and by Michael Russell and Allan Hall,[26,31,32] is catalytic synthesis starting from readily available precursors (CO_2, CO, CH_4, H_2, NH_3, H_2S etc.), on the surface of metal sulphides, particularly iron–nickel sulphides. These are deposited around hydrothermal vents which emit hot water saturated with dissolved minerals, on the ocean floor. A particular advantage of this location for the origin of life is that it would be protected from the potentially lethal ultraviolet radiation to which the early Earth was exposed. *In vitro* synthesis of amino acids and other organic molecules from simpler precursors, catalysed by metal sulphides, has been demonstrated in the laboratory.

Thus we can confidently assume that not only did the post-cometary Earth have a wide selection of bio-relevant molecules supplied from space, but it also had its own mechanisms for generating more.

The first living cells

The best evidence for earlier forms of life is their remains, preserved in recognizable form as fossils in sedimentary rocks. Unfortunately, unaltered sedimentary rocks are never older than about 600 million years. More ancient rocks have been transformed to varying extents by heat and pressure, and are more properly classified as metamorphic. Since we have, on various grounds, good reason to suppose that life began more than 2 bya, the chances of finding clearly recognizable fossils of the earliest life forms are clearly not good. Living organisms, even when they leave no visible trace, can however leave their unique chemical signature in the rock ('molecular fossils'), e.g. in the form of a change in the ratio of the two stable carbon isotopes ^{12}C and ^{13}C, or the presence of characteristic hydrocarbons derived from microbial lipids. Suggestive, if not conclusive,

evidence for very early life has come from measurements of the $^{12}C/^{13}C$ ratio in 3.85 billion-year-old metamorphic rock from Akilia Island in Greenland. Living organisms favour the lighter isotope when fixing CO_2, and so have higher values of the ratio than is found in carbon of non-biological origin. Stephen Mojzsis and coworkers[28] found the ratio in the Akilia rock to be significantly higher than expected in the absence of living material: they took this to be a signature of life. Japanese geologists Yuichiro Ueno and Shigenori Maruyama[37] found organic matter in 3.5 billion-year-old rock from the Pilbara region in north-western Australia which, after analysing the carbon, they concluded was of biological origin. Under the microscope, the rock could be seen to contain tiny fluid inclusions with gas bubbles inside. By careful extraction of the bubbles the geologists could show that they contained methane with a carbon isotopic composition indicative of a microbial origin.

While we are unlikely to find clearly recognizable individual cellular fossils dating from these earliest times, we might hope to be able to recognize the fossilized remains of substantial masses of microbial material if such things occur. In fact they do, in the form of *stromatolites*, which are layered mats of photosynthetic cyanobacteria (Fig. 2.2), together with other micro-organisms. These form, even today, in certain very shallow tropical marine environments. We discuss these in more detail later, but their significance for our present purposes is that recognizable fossilized stromatolites are common in the geological record. The oldest example so far has been found by Abigail Allwood, Malcolm Walter and others in the 3.43 billion-year-old Strelley Pool Chert of the Pilbara region in Western Australia. Chert is a sedimentary rock with a greater than average ability to resist metamorphism, and at various sites around the Earth has provided our oldest fossils. The Strelley Pool Chert contains many laminated carbonate structures which Allwood, Walter et al. convincingly interpret as being of biological, presumptively stromatolite, origin.[1,2] In addition to the structural evidence they found a 250-fold enrichment of rare earth elements in the carbonate laminae, which is a chemical signature of microbial origin.

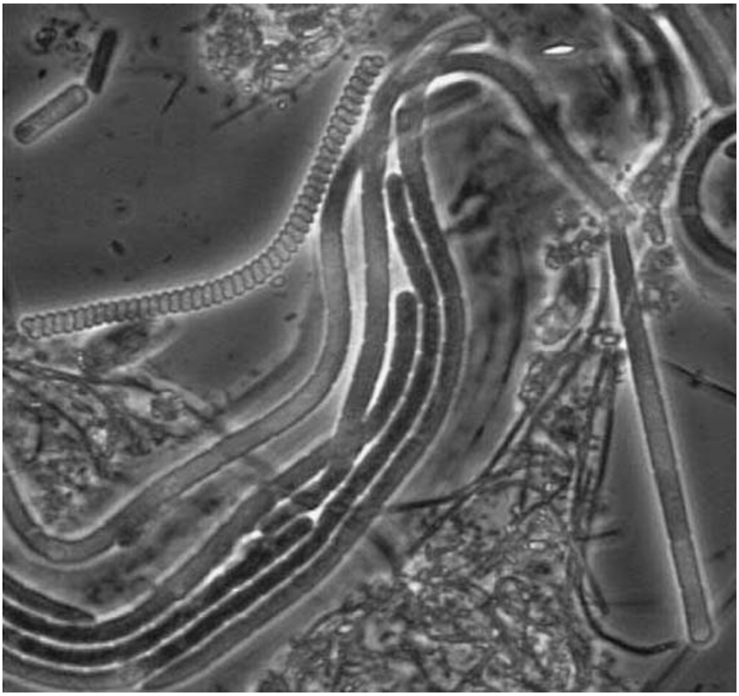


Fig. 2.2 Mixture of cyanobacteria from microbial mat community.
Source: Courtesy Rolf Schauder, University of Frankfurt; Mark Schneegurt, Wichita State University; Cyanosite, www-cyanosite.bio.purdue.edu.

Thus we now have evidence for life, in the form of cyanobacteria 3.43 mya, methanogenic (methane producing) microbes 3.5 bya, and suggestive evidence for some form of microbial life 3.85 billion years before the present day.

There appears to have been a rather short period of time, measured on a cosmic scale, between that point in the Cometary Era when the sea, although still hot, was not too hot for at least thermophilic life to exist, and the first appearance of living cells. In this period, out of the dilute soup of prebiotic organic molecules present in the sea, originating from space and synthesized in terrestrial lightning or on the ocean

floor, there somehow emerged the first self-replicating living cell. It was separated from the outer aqueous environment by a semipermeable membrane, and contained within itself all the biochemical machinery for taking up nutrients from the water, generating energy for biosynthesis, using that energy to build its own proteins, lipids (fatty substances) and nucleic acids from smaller precursors, and then to divide into new daughter cells. Each daughter cell was provided with a long strand of DNA, the blueprint for new cell manufacture. How did this come about?

Without equivocation, let us acknowledge that we do not know. It seems unlikely that we will ever know. We cannot expect to find fossils where there were no structures. What we can do is outline certain criteria which any pre-cellular chemical pathway of evolution must satisfy. Living cells consist largely of polymers, very long molecules consisting of strings of subunits. In any given biological polymer molecule the subunits are not identical but are of the same general chemical nature. Proteins are built up of 20 different kinds of amino acid, ribonucleic acid (RNA) is built up from 4 kinds of nucleotide, deoxyribonucleic acid (DNA) from 4 kinds of deoxynucleotide, and polysaccharides from a variety of different monosaccharide (sugar) units. To create any of these polymer molecules from its subunits, energy must be supplied. In the present-day living cell, segregated from the outside world by its membrane, there are complex biochemical mechanisms for creating high-energy forms of the subunits then combining them under the control of specific enzymes, and (in the case of RNA, DNA, protein) specifying the sequence along a linear template, to give rise to the final polymer. Challenges to our understanding of pre-cellular evolution include coming up with plausible chemical mechanisms, starting from the simple molecules supplied by the comets in the primeval ocean, for the formation of high-energy forms of the various biological polymer subunits. We also need mechanisms much simpler than the present-day complex, highly coordinated intracellular biochemical

replication apparatus, for laying down the subunits in some appropriate sequence. The evolution of photosynthetic and metabolic energy-generating systems, and of the enzymic machinery for complex biosynthetic pathways, also needs to be addressed. This is a field where speculation is rife and facts are few: I do not propose to add to it here.

I will, however, mention that there is a school of thought according to which, before the DNA-RNA-protein life-forms that exist today, there was a simpler life-form based primarily on RNA. The genetic information resided in the nucleotide sequence of the RNA molecules and chemical change was brought about by the catalytic properties of RNA[18]. Short of finding some extraterrestrial planet where life has evolved just to this stage, it is hard to see how this, and competing theories, can ever be put to the test.

Interestingly, the last stage in pre-cellular evolution – the encapsulation of some primitive biochemical system within a surrounding membrane – may be the least problematic. Certain fatty substances such as fatty acids (Fig. 2.1), which may well have been present in the primeval ocean from the cometary input, have a natural tendency to form thin sheets, called *lipid bilayers* (Fig. 2.3a). In a lipid bilayer the lipid molecules are arrayed side by side with their hydrophobic (water-hating) hydrocarbon chains forming the interior of the bilayer and their hydrophilic (water-loving) groups on the exterior surface in contact with the surrounding aqueous medium. These lipid bilayer sheets can turn back on themselves to form globules, enclosing some of the surrounding liquid (Fig. 2.3b).

While acknowledging that we have very little idea how it happened, let us take the pre-cellular stage of biological evolution as given and outline the rest of the story, about which we do have quite a lot of information from the fossil record. Present-day life-forms can be broadly divided into two categories in accordance with how their genetic material is arranged within the cell. In *eukaryotes* – higher organisms, including ourselves – the genetic material exists in *chromosomes* (complexes of DNA with specific proteins) segregated from the rest of the cell by a membrane, within a structure known as the nucleus (Fig. 2.4b). Eukaryotic cells also have their

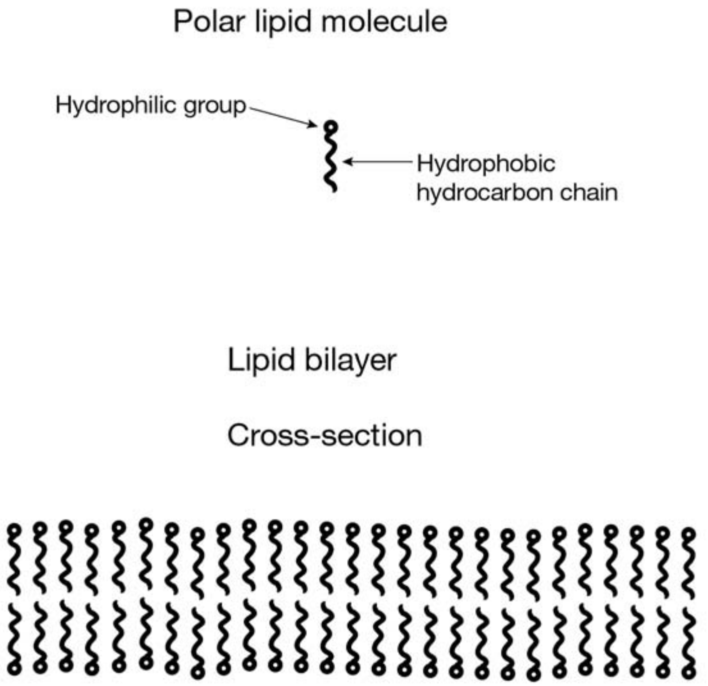


Fig. 2.3 Encapsulation of aqueous medium within a lipid bilayer membrane. (a) Lipid bilayer, the structural principle of all biological membranes. When polar lipid molecules are suspended in water they spontaneously aggregate to form a double layer with the hydrophilic polar groups in contact with water at the upper and lower surface, and the hydrocarbon chains side-by-side forming the hydrophobic interior.

respiratory energy-generating enzymes segregated within separate membrane-bounded structures (the mitochondria), and if they are plant cells they have their photosynthetic energy-generating system also segregated within separate membrane-bounded structures (the chloroplasts). In *prokaryotes* – the bacteria – there is no segregation of function within the cell. The DNA, the respiratory energy-generating enzymes, the photosynthetic apparatus (in the cyanobacteria and other photosynthetic forms) all lie freely exposed in the cytoplasm (the internal medium) of the cell (Fig. 2.4a).

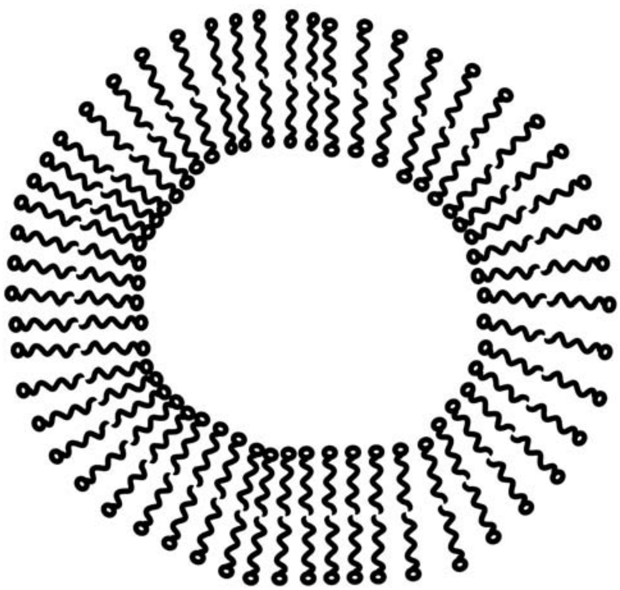


Fig. 2.3 Encapsulation of aqueous medium within a lipid bilayer membrane. (b) Lipid bilayer globule, or 'liposome'. A bilayer sheet can round up and join up with itself in this manner to form a globule or sac, enclosing some of the surrounding aqueous medium.

The very first cells were undoubtedly prokaryotes: simple membrane-bounded sacs, filled with enzymes for energy generation, biosynthesis and replication. The prokaryotes are divided into two kingdoms: the *bacteria*, by far the most common, and the *archaea*, sometimes distinguished as the *eubacteria* and the *archaebacteria*. They are both very ancient. There are a number of subtle biochemical differences between them.

Although the earliest prokaryote cells to show up in the fossil record, in the 3.43 billion-year-old fossilized stromatolites from Western Australia, look very like cyanobacteria – simple photosynthetic algal cells which are still common and highly successful today – we cannot assume that the very first cells, most probably preceding those of which we have fossils, were photosynthetic. Indeed it seems much more likely that they were *heterotrophic* and/or *chemoautotrophic*. Heterotrophic microbes live by consuming pre-existing organic compounds: the presumptive early forms

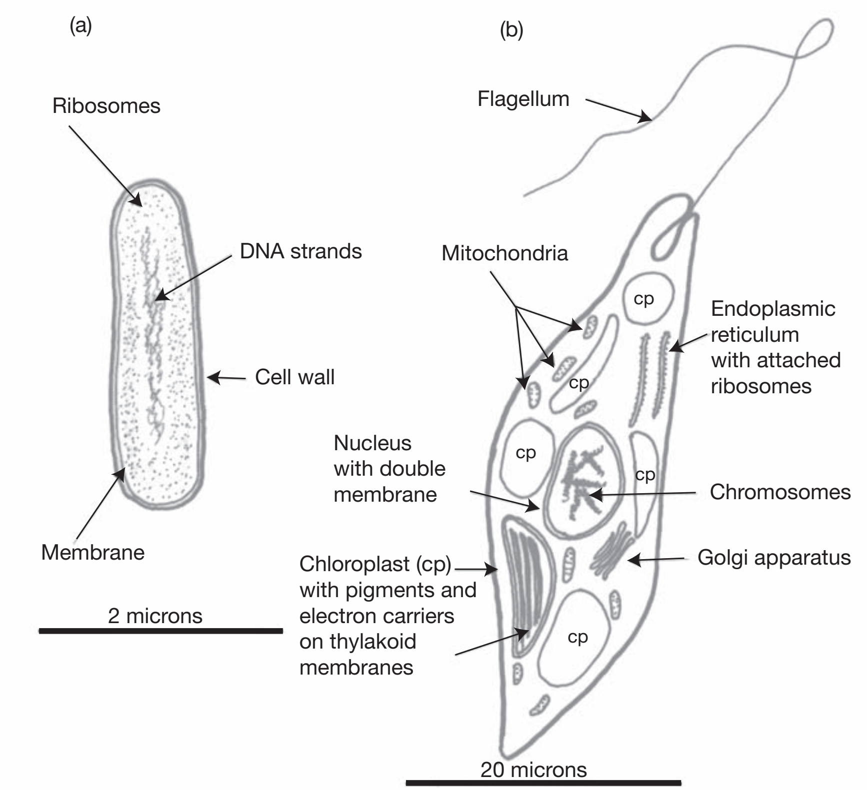


Fig. 2.4 Prokaryote and eukaryote cell structure. (a) Typical bacterial cell. (b) Cell structure in a unicellular eukaryote alga in the class *Euglenophyta*.

would have utilized the various substances dissolved in the primeval ocean. Once self-replicating heterotrophic prokaryotes had arisen, however, they would multiply until the compounds upon which they lived became depleted. Chemoautotrophs, by contrast, would have had a reliable food supply. The general principle upon which these microbes work is to transfer hydrogen from highly reduced substances such as H_2, H_2S, CH_4, NH_3, to hydrogen acceptors such as CO_2. The overall process can be represented as:

$$4H + CO_2 \rightarrow (CH_2O)_{\text{cellular material}} + H_2O$$

The hydrogen transfer is accompanied by the generation of biochemical energy in the form of ATP (adenosine triphosphate), which can then be used to bring about the synthesis of all the compounds the cell needs for growth. Hydrogen sources such as those listed above, and a hydrogen acceptor in the form of CO_2, were plentiful in the early Earth. The hydrothermal vents on the ocean floor are plausible locations for the evolution of the first chemoautotrophs.[26,31]

Although it seems likely that the very first organisms were chemoautotrophs and heterotrophs, the presence of 3.43 billion-year-old stromatolites in the fossil record tells us that a new mode of energy generation – in the form of *photosynthesis* – appeared very early in the history of life on Earth. Whether this was by chance just once in one particular cell, or many times in many cells, we cannot say.

The Prokaryotic Photosynthesisers

While we know a great deal about how photosynthesis works in present-day plant cells, the form it took in the very first photosynthetic cell can only be a matter for conjecture. What we can say is that this cell must have contained molecules of a *pigment* – a substance whose chemical structure is such that it absorbs light in the visible region of the spectrum. Assuming the fundamental principle was the same as that which operates in present-day photosynthetic organisms, a pigment molecule in its excited state after absorption of solar radiant energy would have transferred one of its electrons to another molecule (i.e. it would have chemically 'reduced it') (Fig. 2.5). In this way 'chemical reducing power' would have been generated, which by unknown pathways (although we know the pathway which exists today) was used to convert carbon dioxide, the form of carbon present in abundance, to reduced (hydrogen-containing) compounds such as sugars which could be metabolized in the usual way.

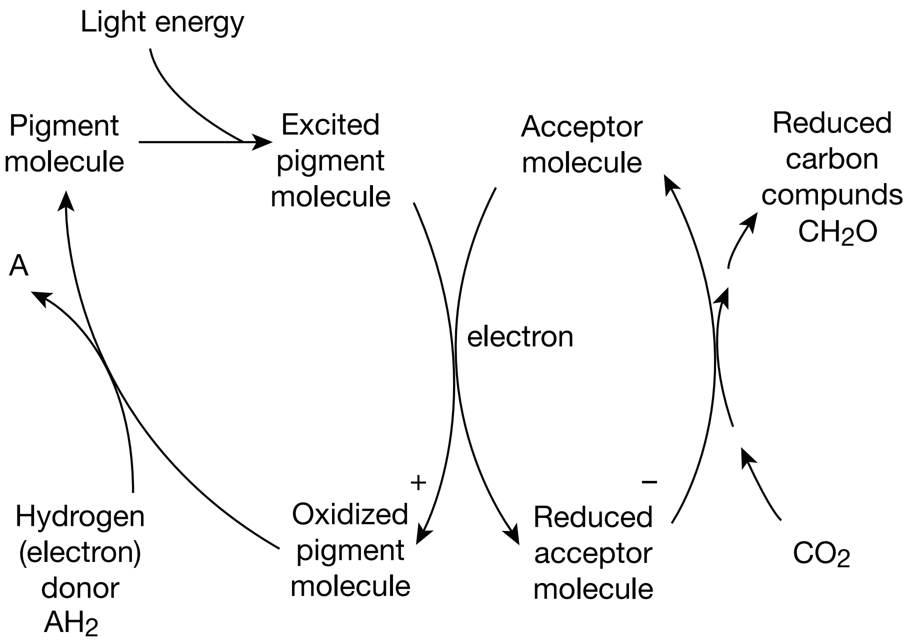


Fig. 2.5 Outline of the overall photosynthetic process. The pigment molecule is raised to an excited state by its absorbed light energy, and transfers one of its electrons to an acceptor molecule. The pool of reduced acceptor molecules created in this way provides the hydrogen which, by a process involving numerous biochemical steps, reduces CO_2 to carbohydrate and other metabolizable carbon compounds. The highly reactive, positively-charged pigment molecule regains neutrality by removing an electron (equivalent to a hydrogen atom) from a hydrogen donor such as H_2S, NH_3 or H_2O. In the latter case, molecular oxygen is liberated as a consequence.

On losing an electron, the pigment molecule was left with a positive charge and became a very powerful oxidizing agent with the ability to pull an electron or hydrogen atom out of other molecules and thus restore itself to electrical neutrality. The positively charged pigment molecules may have initially used readily oxidized, although not always very plentiful, substrates such as hydrogen sulphide (H_2S) and ammonia (NH_3), but we know that eventually there arose forms which – like modern algae and higher plants – could pull hydrogen atoms out of a hydrogen donor present in unlimited supply, namely water itself. This may have seemed like a

good idea at the time but, as it was accompanied by the liberation of free and highly reactive molecular oxygen, it was to have enormous consequences for the future pathway of biological evolution. What had arisen in these very first oxygen-yielding photosynthetic cells was no doubt something much simpler than the modern coordinated multi-component photosynthetic system, and working probably much less efficiently, but with a similar overall effect.

We do not know why pigmented cells had arisen in the first place. The survival of any such cells that arose by chance may have been favoured by the protection that the pigment would provide against the damaging effects of intense solar radiation, especially the ultraviolet component. The pigments, by absorbing radiant energy, would also tend to increase the temperature of the cell and thus accelerate its metabolism. Whatever the original advantage of the pigments, the day came when the biochemical environment within the cell was such that the excitation energy of the molecules was able to bring about useful chemical change instead of merely dissipating as heat.

The next great stage of biological evolution – the Age of the Prokaryotic Photosynthesisers – had begun. From our hypothetical first photosynthetic bacteria, with their no doubt somewhat haphazard mechanism for capturing radiant energy and turning it into chemical energy, there had already arisen, at least as long as 3.5 bya (towards the end of the Earth's first billion years of existence) the first cyanobacteria, the photosynthetic prokaryotes which still exist and thrive today. These were the dominant form of life for about the next 2 billion years.

Because of their minute and delicate nature the prokaryotic cells of what is now called the Archaean era have, with one exception, left little trace in the fossil record. The exception is the type of filamentous cyanobacteria that formed stromatolites. These are large laminated colonies in which layers of cyanobacteria, silt and limestone are formed, building up one layer above the other, until they eventually form stony columns with rounded heads, perhaps 50 cm high (Fig. 2.6). Living stro-

Fig. 2.6 Stromatolites in Shark Bay, Western Australia.
Source: Photo by Kevin Crane, courtesy CALM, Shark Bay, WA.

matolites are found today in shallow, warm, highly saline waters such as Shark Bay in Western Australia, and certain parts of the Caribbean. They thrived in the conditions existing in the Archaean sea and may have been the dominant form of cyanobacteria. Their closely adherent coating of fine mineral particles means that stromatolites have survived well in the fossil record.

In addition to the photosynthetic micro-organisms, which used light energy to synthesize cellular material out of CO_2 and mineral nutrients, there must also have existed *saprophytic* bacteria which used dead cyanobacteria as a food source, reconverting organic matter to CO_2 and inorganic materials. There may well also have existed predatory micro-organisms which used living cyanobacteria as a food source. Thus the first food chain came into existence, with cyanobacteria forming the base. It is important to realize that the decomposers played an essential role in this early ecosystem: without them, cyanobacterial proliferation would have continued

until it came to a dead stop, when the most limiting nutrient, probably nitrogen, was all locked up in biomass. Evolution needed continual turnover.

OXYGENATION OF THE OCEAN AND ATMOSPHERE

The reaction at the heart of photosynthesis is the splitting of the water molecule by means of the light energy harvested by pigments. The hydrogen atoms are transferred to specific chemical carriers which, in the presence of the appropriate enzymes, achieve the reduction of CO_2 to carbohydrate, which can then be used by the cell for all its metabolic and biosynthetic purposes. The oxygen atoms left behind by removal of hydrogen from water combine to give molecular oxygen. Thus from the first appearance of photosynthetic prokaryotes, just over 3.5 bya, oxygen began to be liberated into the Earth's ocean and atmosphere.

At that time the planet was very much in a chemically reduced state and for a long time all the liberated oxygen was mopped up by the various reducing substances – H_2, CH_4, H_2S, various organic compounds, ferrous iron in rocks, etc. – present in the ocean, atmosphere and mineral crust. Thus, for a billion years or so, the Earth's surface continued to provide a highly anaerobic environment for the life-forms that had so far developed. All this time, however, the planet's store of reducing materials was being slowly but inexorably used up, until about 2.4 bya, there were no reductants left and free molecular oxygen began to accumulate in the atmosphere and the ocean. By 2 bya it had risen to 3% and it continued to climb steadily, reaching its present level of ~21% about 600 mya (Fig. 2.7).

When oxygen first appeared it would have been toxic to most of the organisms then living, just as it is to anaerobic organisms today. But soon, new prokaryotic organisms appeared which not only tolerated oxygen but could use it to carry out oxidative metabolism of their carbohydrate, a much more efficient method of biochemical energy generation than the

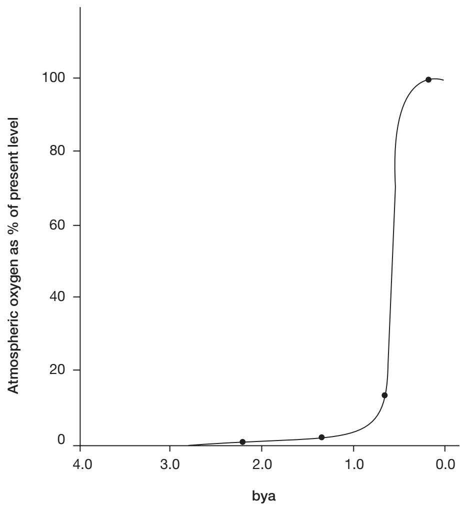


Fig. 2.7 Oxygen content of the atmosphere, from the early prebiotic Earth to the present day.
Source: Graph based on data in 'Atmosphere', *Encyclopedia Britannica*, 15th edn (1993), vol. 14, pp. 320–321.

purely anaerobic metabolic pathways. This meant that the non-photosynthetic part of the biota, the saprophytes and predators, were given a considerable boost. A quite different, but highly significant, additional consequence of the appearance of free oxygen was that *ozone* – the triatomic form of oxygen (O_3) – began to form, giving rise to what we now call the *ozone layer* in the upper atmosphere. Ozone has an absorption peak in exactly that part of the UV spectrum, wavelength ~260 nm, where the nucleic acids of living organisms also absorb UV (Fig. 2.8). The ozone layer thus came to form a protective shield around the planet, filtering out the dangerous UV photons with their potentially lethal effects on all DNA-based life-forms. It became possible for organisms to survive in the upper layers of the ocean where solar radiation was intense. Furthermore, one of the barriers to eventual colonization of the land had been removed.

The appearance of oxygen in the atmosphere also had implications for the Earth's temperature. The luminosity of the Sun when life began, ~3.5–4 bya, was about 25% lower than it is today and the greenhouse effect

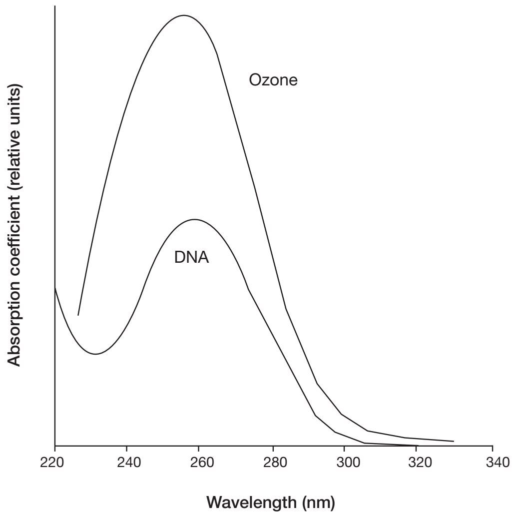


Fig. 2.8 Ultraviolet absorption spectra of ozone and of DNA. We can see that the molecules of gaseous ozone high in the stratosphere have an absorption band centred on 260 nm. This is very well positioned to remove the wavelengths of solar UV which would otherwise be absorbed by, and cause damage to, the DNA which is the genetic material of living organisms.

of the gases in the atmosphere played a crucial role in preventing the oceans from freezing over. Analysis of the $^{12}C/^{13}C$ ratio in the organic matter in a ~2.8 billion year-old *palaeosol* (fossilized soil) at Mount Roe in Western Australia[33] indicated that it was formed by *methanotrophs* (methane-consuming bacteria), which in turn implies the presence of high levels of methane (a very potent greenhouse gas) in the atmosphere. As photosynthetically produced oxygen was emitted from the ocean into the atmosphere, it oxidized the methane so that the concentration of this gas gradually fell. The Earth consequently cooled and this may have been the cause of the first known glaciation (the *Huronian* glaciation) around 2.3 bya.[19] Oxygen levels in the ocean were still very low, and it is thought that methane continued to be emitted into the atmosphere at a rate sufficient to prevent total freezing.[30] When atmospheric oxygen began to climb towards its present level, towards the end of the Proterozoic eon (Fig. 2.7), methane concentration became vanishingly small and the Earth cooled,

leading to at least two and possibly four ice ages. Palaeontologist Andrew Knoll suggests[22] that the first was a bit earlier than 765 mya and possibly confined to Africa; the second and third, at ~710 and ~600 mya, were global in extent. The last was relatively minor and occurred before the Cambrian (commencing 543 mya) period began. More controversial is the *Snowball Earth* hypothesis, originally proposed by Caltech geologist Joe Kirschvink in 1989, that the Earth has frozen over completely three times: the 'Makganyene', 'Sturtian' and 'Marinoan' glaciations at 2.22 bya, 710 mya and 640 mya respectively. It is claimed that during these episodes ice sheets covered the land and sea ice covered the oceans. The total freeze-over was terminated each time by atmospheric warming resulting from accumulation of CO_2 discharged into the atmosphere by volcanoes. The exact extent of these early glaciations in Earth's history remains a matter of vigorous debate.

APPEARANCE OF THE FIRST EUKARYOTES

Some time between 2 bya and 3 bya, after the appearance of the first photosynthetic prokaryotes, there was an evolutionary event of enormous significance. This was the formation of the first eukaryotic cells, with their genetic material organized into chromosomes segregated within a membrane-bounded nucleus, and some of their metabolic functions carried out within distinct organelles: mitochondria for oxidative metabolism and chloroplasts (in plant cells) for photosynthesis (Fig. 2.4b). The earliest recognizably eukaryotic fossils, of protozoan-like organisms, are 1.8–1.9 billion years old, found in rocks in China.[21] What appear to be fossils of eukaryotic protozoa have been found in 1.4–1.5 billion-year-old carbonaceous shales in northern Australia.[17] Molecular fossils in the form of *steranes*, hydrocarbons derived from cholesterol-related sterols of a type believed to be synthesized only by eukaryotes, have been found in 2.7 billion-year-old shales from the late Archaean era, in north-western

Australia.[4,5] It seems likely that eukaryotes first appeared more than 2.7 bya. The biosphere was still dominated by cyanobacteria and the other prokaryotes dependent on them, and indeed remained so dominated for perhaps another 1½ billion years or so. But the stage was now set for the next great era of biological evolution – the Age of the Eukaryotes.

The most important distinction between the eukaryotes and the prokaryotes is the great difference in the way the genetic material is organized. In the prokaryotes the DNA exists as a very long strand, joined together at the ends to form a loop, which lies haphazardly in a loose tangle in the cytoplasm of the cell. In the eukaryotes the genetic material is segregated within a separate structure, the *nucleus,* bounded by a double membrane. Within the nucleus the DNA is complexed with certain specific proteins, known as *histones,* and forms discrete structures known as *chromosomes.* These typically occur in pairs and there are usually several pairs of chromosomes in each nucleus. There is a mechanism within the eukaryotic nucleus which ensures that on nuclear division, prior to cell division, each daughter nucleus receives one copy of each chromosome. The higher level of complexity and coordination that exists in the eukaryote nucleus is likely to reflect the need for much greater control of the gene activity in the structurally and functionally more complex eukaryotic organisms. The establishment of precise genetic control mechanisms at this early stage paved the way for the appearance, much later, of multicellular organisms whose complexities of structure and function were vastly greater than those of their unicellular ancestors.

How the eukaryote nucleus evolved remains a mystery. Details of intracellular structure are not preserved in fossils, so we have no idea what early versions of the nucleus in the first prokaryotes to go down this pathway looked like. One hypothesis (out of many), which takes account of the fact that eukaryotes have inherited aspects of their biochemistry from both bacteria and archaea, is that the first stage in eukaryote formation was a close association – both physical and symbiotic – between a eubacterium and an archaebacterium.[26] The eubacterium went on to become established within the archaebacterium as an intracellular symbiont, and even-

tually took control of the combined genetic material and evolved into the nucleus.

For two of the other components of the eukaryotic cell, the mitochondria and (in plants) the chloroplasts, a very plausible supposition (overwhelming in the case of chloroplasts) is that they originated by capture and symbiosis. We have already hypothesized that certain early prokaryotes were predatory, i.e. they lived by consuming other prokaryotic cells. This being so, it seems likely that some early eukaryotes were also predatory. Some of these no doubt operated by engulfing their prey, like the amoeba does today. It is now thought likely that occasionally the engulfed cells were not digested but continued to multiply within their hosts, and over millions of years settled down to a stable and mutually beneficial relationship that evolved into the organelles – mitochondria and chloroplasts – that we see today. A full account of this evolutionary hypothesis can be found in various books by microbiologist Lynn Margulis.[24]

The *organelles as symbionts* hypothesis was first mooted (in 1905 and 1907) by two Russian cytologists, Mereschkowski and Famintzin, who proposed that chloroplasts were descended from symbiotic cyanobacteria. This has a lot of supporting evidence and is now widely accepted. Both chloroplasts and mitochondria contain DNA, lying free in the interior just as in prokaryotic cells, and have the enzymes for synthesizing RNA using the DNA as a template. Both have protein-synthesizing systems (ribosomes and associated biochemical machinery) specifically of a prokaryotic rather than a eukaryotic type. In chloroplasts, the pigments (chlorophyll etc.) which harvest light for photosynthesis are present in membranes arranged in parallel layers as they are in cyanobacteria. One of the most persuasive arguments in favour of the organelle-as-symbiont hypothesis is the fact that many examples of symbiotic relationships between algae (even eukaryotic species) or bacteria, and host cells within which they reside, are known in the modern world. Unlike many of the modern-day symbionts, neither chloroplasts nor mitochondria can live separately from their host cells today, so we must suppose that over the 2–3 billion years since they were

first established within other cells, they have relinquished some of their essential biosynthetic and metabolic functions to the host. The benefit to the descendants of the original host cells has been immense. The putative original cyanobacterial symbiont conferred the ability to live photosynthetically. The prokaryotic symbiont that was to evolve into mitochondria may well have brought with it a highly efficient enzyme system for oxidative metabolism. Since the biosphere as a whole was still highly anaerobic at the time (> 2.7 bya) eukaryotes first evolved, we have the interesting possibility that mitochondria – whose essential function is oxidative metabolism – first appeared within plant cells, where they would have immediate access to the photosynthetically liberated oxygen before it diffused into the outer world and was immediately mopped up by the omnipresent chemical reductants.

Over the next 1.5–2 billion years the eukaryotic cells evolved into innumerable different forms occupying a variety of ecological niches. They developed a much wider range of morphological and ecological characteristics than their prokaryote relatives, but like them remained single-celled organisms. Both prokaryotes and eukaryotes can form filaments, with chains of cells joined together, but each cell in the filament is essentially identical and performs the same functions, and such filamentous species are still regarded as single-celled forms.

THE FIRST MULTICELLULAR ORGANISMS

About 1 bya the next great event in the evolutionary story occurred – the appearance of multicellular organisms. This is thought to have happened first among the algae. Fossils of a simple multicellular red alga have been found in Canada in rocks with an estimated age of 950–1260 million years.[21] Fossil evidence for the first multicellular animals (heterotrophic nonphotosynthetic multicellular eukaryotes) dates from about 580 mya[29] in the Ediacaran period, which extends from the end of the Marinoan

glaciation (the last global, possibly 'Snowball', glaciation) at 630 mya to the beginning of the Cambrian period at 542 mya. It was in the following Cambrian period (542–488 mya), however, that the first massive proliferation of animal lifeforms – the *Cambrian Explosion* – took place.

In a multicellular organism the different cells, although genetically identical, can develop in different ways so as to provide a range of specialized functions: photosynthesis, metabolite transport, nutrient extraction etc. in plants; digestion, locomotion, vision, nervous control etc. in animals. In every such organism there is, in addition to all the structural genes for making the actual proteins etc. of which it is composed, a suite of developmental genes which precisely control its growth from a fertilized egg into the kind of animal or plant it is meant to be.[6] A particularly important set of these, the *Hox* genes, are found in all bilateral animals – insects, reptiles, mammals etc. – and appear to have originated as far back as the Cambrian period.[6,22]

In addition to being much more complex and versatile, multicellular organisms are generally bigger than unicells, and size brings its own advantages. A disadvantage of increased size, however, is that oxygen molecules from the surrounding seawater have further to travel to reach cells not on the surface of the organism. The palaeontologist, Richard Fortey, suggests that the relatively late proliferation of multicellular organisms may have been because they could not function efficiently until the atmospheric oxygen concentration reached a high level.[13] The present value of ~20% was not reached until something less than 1 bya.

Once multicellularity had found a firm toehold on the evolutionary ladder, these organisms proliferated, and radiated into an innumerable variety of forms. From here on in the story of Life on Earth we have a wealth of information in the fossil record, especially of organisms with hard shells, skeletons or other resistant structures. Since my concern in this chapter is the fundamentals of what happened in evolution rather than the specifics, and mainly the implications for humankind, we shall not trace the development from the first well-defined multicellular organisms –

the Ediacaran fauna of ~580 mya – to the enormously rich biological diversity that exists today, in any detail. The books by Richard Fortey,[13] Andrew Knoll[22] and Richard Dawkins[10] or the numerous contributions by Stephen Jay Gould will give a more complete picture.

EVOLUTION OF HUMANKIND – THE AQUATIC PHASE

Among all the millions of life-forms that now inhabit our planet, we shall give particular attention to the evolution of just one – humankind, *Homo sapiens* (Fig. 2.9 a–e). What are we? We began as quarks, electrons and gluons in that unthinkably concentrated pinpoint of energy which we call the 'Big Bang', 13.7 bya. After ~7½ billion years we had become stardust: atoms synthesized in the furnace of innumerable suns, blown out into space by fiery supernovae, aggregated into dust particles in giant clouds, settled out into the accretion disk around a newly created star, and finally deposited as cometary debris on the surface of one of the planets forming within that accretion disk.

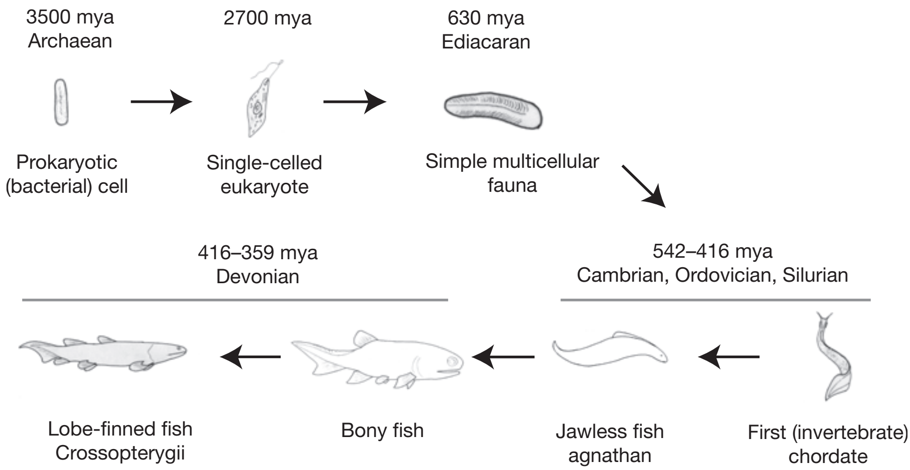
We are, of course, eukaryotes, parting company with our prokaryote ancestors 2½–3 bya when we acquired chromosomes within a nucleus, and intracellular organelles. We are animals not plants, acquiring our energy and nutrients for growth by consuming other organisms, descended perhaps from some predatory protist gobbling up the other cells it encountered in the warm, increasingly oxygenated seas of the Proterozoic eon. We are multicellular, leaving our unicellular relatives behind perhaps ~0.6 bya in the Ediacaran period. Among the 23 or so major divisions, or *phyla* (singular *phylum*), of the animal kingdom we are *chordates*, i.e. members of the phylum *Chordata*, defined by possession of a nervous system organized around a nerve chord, closely adjacent to a stiff supporting rod, running from top to tail down our back. This crucially important organizational principle which differentiates chordates from other phyla such as the molluscs and the arthropods first appeared in the Cambrian period,

~542–488 mya, probably in the form of minute unassuming invertebrate creatures such as *Pikaia* found in the Burgess Shale of Canada, a famously rich source of Cambrian fossils.[16] From these there developed small and primitive fish precursors without backbones or functional jaws, which came on the scene during the Cambrian, but flourished in the oceans of the Ordovician (488–444 mya) and Silurian (444–416 mya) periods (Fig. 2.9a).

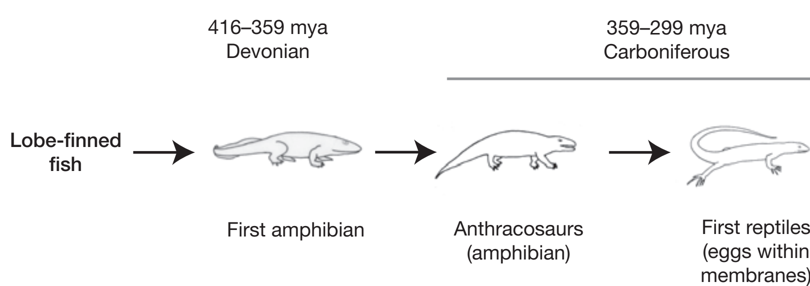
In the Devonian period (416–359 mya) the first fish of modern type, with backbones and jaws, appeared. As the first animals with backbones, these are of particular significance as the ancestors not only of humankind, but of all vertebrate animals. One especially significant group was the *Crossopterygii* or *lobe-finned fishes*. These had muscular fins supported by bony elements, and functional lungs. About this time or even earlier, plant life had finally succeeded in establishing itself on land and so there were now terrestrial ecosystems, full of metabolizable organic matter, waiting to be colonized. From some species of lobe-finned fish there developed the first *tetrapods*, animals in which the fins had evolved into four short limbs. One was *Acanthostega*, which lived in the late Devonian period (416–359 mya) in what is now eastern Greenland. Palaeontologist Jennifer Clack, on the basis of her study of the fossil, believes that *Acanthostega* was able to raise its head out of the water, perhaps to breathe oxygen (it had lungs) or to catch prey, but it would not have been able to leave the water.[7] From forms such as *Acanthostega* the first amphibians (Fig. 2.9b) arose towards the end of the Devonian. The stumpy fins had by now developed into crude, no doubt inefficient, legs on which the creatures were able to crawl up onto the mud at the edge of the water in search of food, and back to the water again as need arose.

The earliest amphibians which we know from the fossil record, and which could well be on or close to our ancestral tree, are the *icthyostegalians*. These were quite large (perhaps ~0.5 m), long-tailed, long-jawed creatures with short legs ending in anything from 5 to 8 recognizable toes. They probably spent most of their time in the water – fresh water, not seawater – but were capable of walking on land, where they would have eaten arthro-

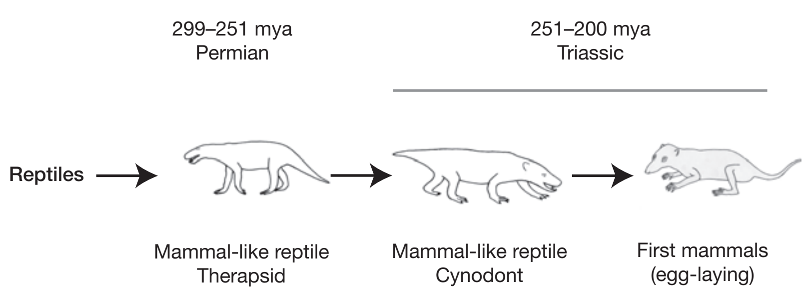
Fig. 2.9 The evolution of humankind.
Source: Simplified representations based on various literature sources, especially *The Marshall Illustrated Encyclopedia of Dinosaurs* (ed D. Palmer, Marshall Publishing: London); *Encyclopedia Britannica*[12]; the plaque on the Pioneer spacecraft, NASA.



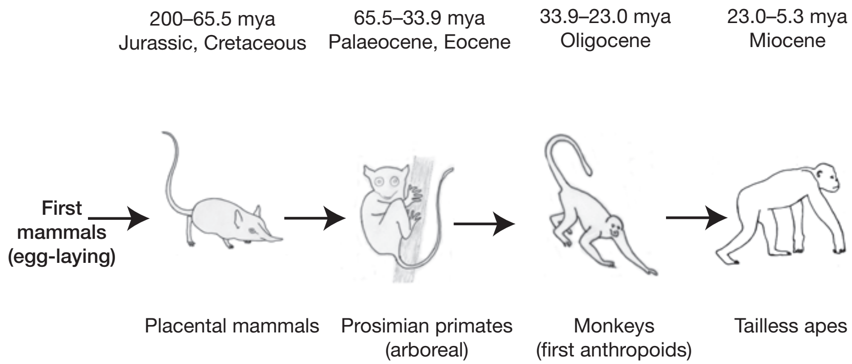
(a) The aquatic phase: from prokaryote cell to lobe-finned fishes.



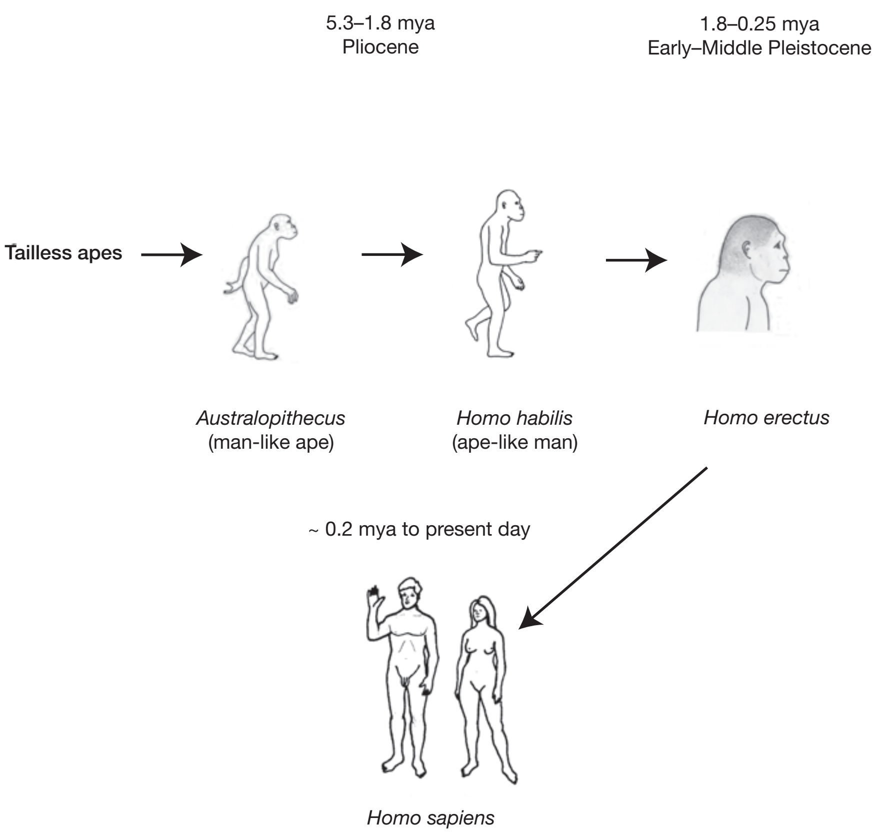
(b) Emergence onto the land: from lobe-finned fish to the first reptiles.



(c) From the reptiles to the appearance of the first mammals.



(d) We take to the trees: from the first mammals to the tailless apes.



(e) The final transition: from apes to *Homo sapiens*.

pods, an animal phylum including insects, mites, spiders, scorpions, which also made the transition from water to land at about this time. Like all amphibians, the icthyostegalians would have returned to the water to breed, just as present-day amphibians such as frogs and newts do.

EVOLUTION OF HUMANKIND – THE REPTILIAN PHASE

In the warm, humid swampy forests of giant ferns and lycopods that flourished in the Carboniferous period (359–299 mya) which followed the Devonian, the amphibians flourished and evolved into a variety of new forms which palaeontologists have sorted into three main groups. One of these groups, the *anthracosaurs,* which included terrestrial as well as aquatic forms, is of particular interest because it was from these amphibians that the *reptiles* – the next great stage in vertebrate evolution – were to evolve. The crucially important development which distinguished the first reptiles from their amphibian ancestors was the formation of a tough membrane around each egg. This protected the eggs from desiccation so that they no longer had to be laid in water, making it possible for reptiles to spread from the swampy forests into drier environments where surface water was not so readily available.

This development arose towards the end of the Carboniferous period and proved of great benefit to this new order of creatures – the *Reptilia* – in the cooler and drier Permian period (299–251 mya) which followed. Although reptiles first appeared in the Carboniferous period, only in the Permian did they become a significant part of the fossil record. The reptiles of the Permian included two of the types, although not of course the actual species, with which we are familiar today – lizards and turtles. From some of these early lizards developed, in the next geological period, the Triassic (251–200 mya), the *dinosaurs,* which dominated the Earth from 225 million to 66 mya.

Fascinating though it is, we shall not concern ourselves here with the further evolution of the dinosaurs because they are not on our family tree. To find our link with the reptiles we must go back to the Permian period.

One particular way in which the reptiles of that time can be classified relates to the openings, in addition to the universally present eye sockets, present in the skull. The presence or absence, and the positioning, of these openings was related to the musculature of the jaw. Turtles had no additional openings. Crocodiles, and those lizards ancestral to the dinosaurs, had two pairs of openings next to the eye sockets: species with this characteristic are generically referred to as *diapsids*. A third group of reptiles, referred to as *synapsids*, had only a single opening, positioned low down on either side of the skull. The ancestor of the mammals is to be found among the synapsid reptiles of the Permian.

These 'mammal-like reptiles' (Fig.2.9c), as palaeontologists call them, flourished and evolved further in the Permian and Triassic periods into both carnivorous and herbivorous types. In the Permian they were probably the dominant reptile form. The earliest mammal-like reptiles were small, but in the Triassic period some species up to 3 m long appeared. The dominance of the synapsid reptiles was, however, about to be challenged because it was in the Triassic that the first dinosaurs appeared, from among their diapsid reptilian cousins. These proliferated through the Triassic period and reached their zenith, in terms of ecological dominance and biological diversity, in the following Jurassic period (200–145 mya).

EVOLUTION OF HUMANKIND – EMERGENCE OF THE MAMMALS

We need waste no sympathy on our synapsid reptilian ancestors for their loss of dominance, because at about the time of the first appearance of the dinosaurs in the Triassic something else of enormous biological significance occurred, namely the development, from some small mammal-like reptile, of the first mammals (Fig. 2.9c). Mammals are characterized by suckling their young with milk, and by having a hairy coat, warm blood and built-in temperature control (*homeothermy*) so that (unlike cold-blooded reptiles) they can function well over a range of ambient temperature.

These characteristics would not have appeared all of a sudden in one new species but must have developed gradually over a long period of time. Although most modern mammal species incubate their young internally in the body, the earliest mammals most probably laid eggs, as their monotreme descendants (the platypus and echidna of Australia) do today.

Despite their early origin and the advantage conferred by homeothermy, the mammals remained an inconspicuous component of the biota throughout the Triassic and the Jurassic periods and right through the following Cretaceous period (136–64 mya). Throughout this time they existed as small animals not unlike their rodent descendants, some living as insectivores or opportunistic scavengers, others as consumers of seeds and roots. They have been described as 'small, secretive and nocturnal'.[14] One interesting consequence of their nocturnal lifestyle was that they had less need of colour vision, and at some time in the Mesozoic era (251–65.5 mya) lost two of the four kinds of cone cells, with their associated visual pigments, in the retinas of their eyes; their reptilian cousins retained them. To this day most mammalian species are *dichromatic* (have only two visual pigments), while birds (dinosaur descendants) and most reptiles have four.[14] Early mammals are very poorly represented in the fossil record and so our knowledge of them is very limited. Thus, up to the end of the Cretaceous period the mammals were not a great success, ecologically speaking, but they were good survivors, having managed to persist for at least 130 million years since their origin in the Triassic.

Their time was coming, because in the Book of Fate it was written that time was up for their dominant rivals, the dinosaurs, at the end of the Cretaceous period. At the transition from the Cretaceous to the Tertiary period (commonly referred to by the experts as the K–T boundary) a great geophysical cataclysm occurred which rendered the planet uninhabitable for the dinosaurs – not permanently, but long enough to ensure that they died out. The exact nature of the event remains uncertain; one suggestion, for example, is massive volcanic eruption, for which there is indeed some evidence. The most widely held view, however, with quite persuasive cumu-

lative evidence, is that Nemesis for the dinosaurs took the form of a large comet, thrown out from the icy regions of the great planets billions of years before and, after orbiting silently through space for countless eons, finally on a direct collision course with Earth. At a particular time on an otherwise unremarkable day, 65.5 mya, a 10 km diameter comet (or asteroid) suddenly appeared out of a clear blue sky and slammed into the Earth (Fig. 2.10) at a place now called Chicxulub, on the Yucatan peninsula on the southern shore of the Gulf of Mexico. The impact created a crater 300 km wide and sent into the atmosphere an enormous cloud of dust consisting of ultrafine particles from the comet itself and finely pulverized rock from the impact site (Fig. 2.11 a, b). When it finally settled, the dust gave rise to a thin layer of clay, 10–40 mm thick, which in suitable geological strata precisely marks the transition from the Cretaceous to the Tertiary periods: Tertiary fossils above, Cretaceous fossils below. The hypothesis that the K–T boundary was caused by a comet/asteroid impact was proposed in 1980 by physicist Luis Alvarez and his geologist son Walter Alvarez. The crucial clue was the totally unexpected finding that the clay layer contains the rare element *iridium* at a concentration 200–300 times that of the adjacent rock layers:[11] this metal is present at vanishingly low

Fig. 2.10 Dinosaur Doomsday – an artist's impression of asteroid impact. Source: NASA.

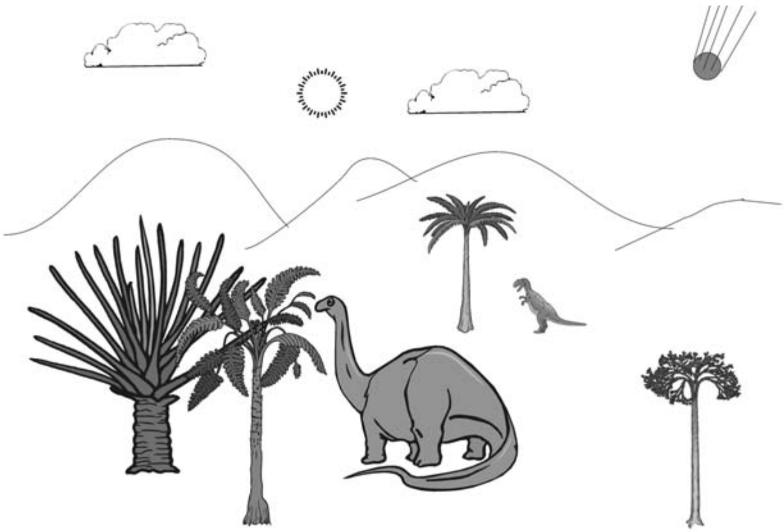
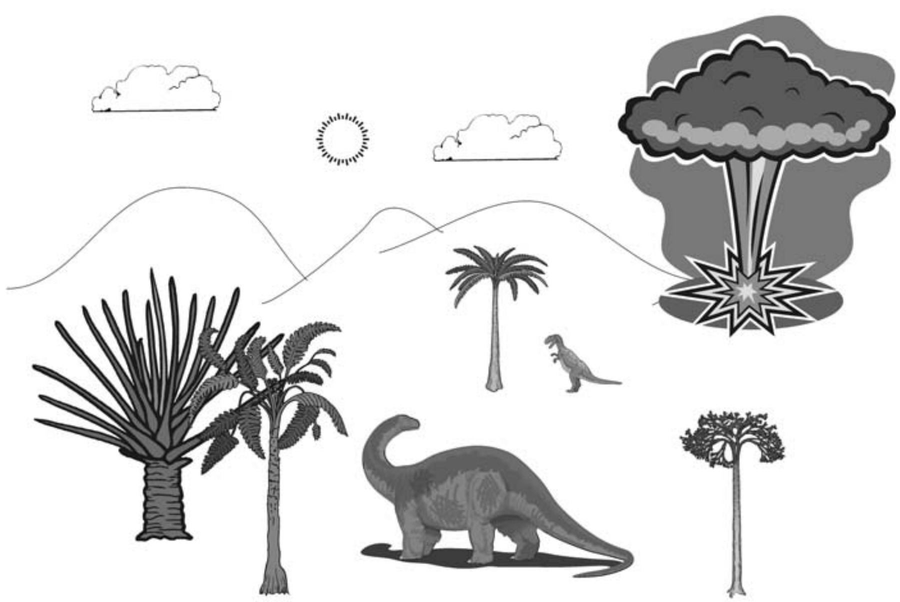
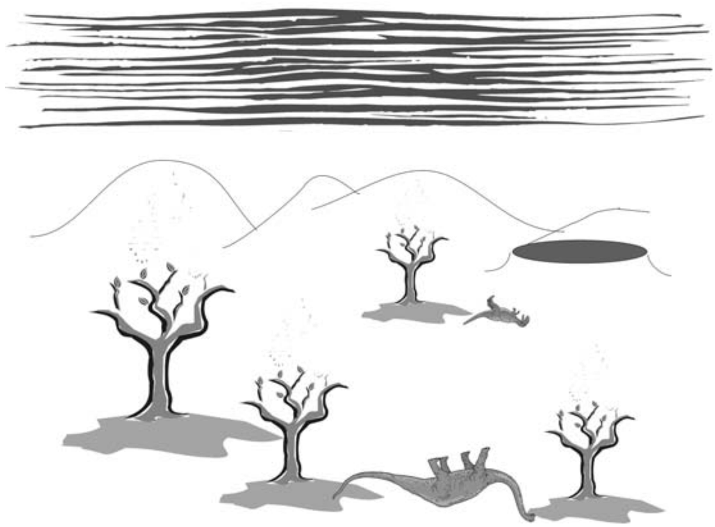


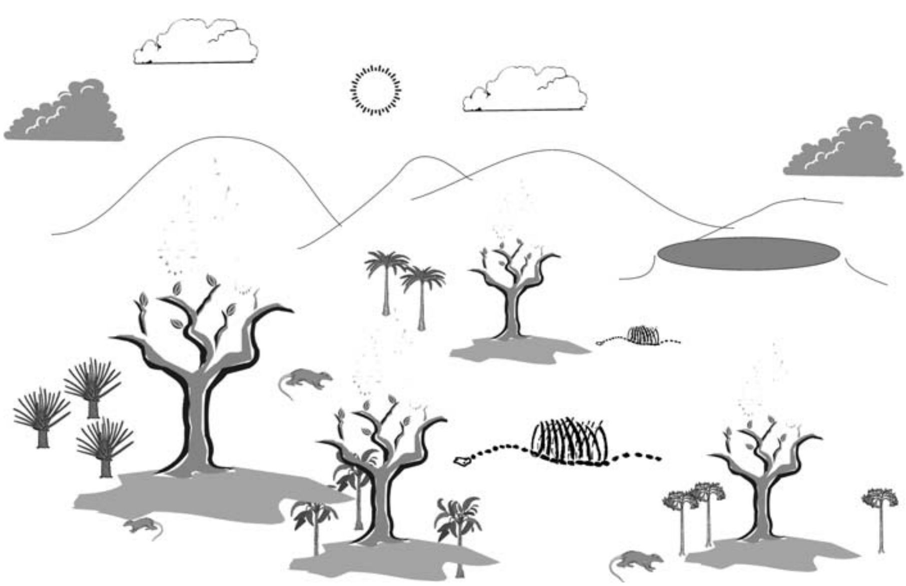
Fig. 2.11 Dinosaur Doomsday. (a) The last day of the Cretaceous period, just before the comet impact.



(b) Just after impact. Enormous quantities of molten and pulverized rock from the impact site and the comet itself are thrown up into the atmosphere.



(c) Darkness covers the Earth. The cloud of comet dust and pulverized rock around the planet cuts off sunlight. The plants cease to grow and the dinosaurs, both herbivorous and carnivorous, die of starvation.



(d) The dust cloud dissipates. New plants begin to grow, but the dinosaurs are dead. Small primitive mammals begin to proliferate. The Age of the Mammals has begun.

levels in the rocks of the Earth's crust, but is relatively plentiful in heavenly missiles such as comets, asteroids and meteorites. Further evidence for the comet theory came from work by the US Geological Survey which revealed the presence of a special kind of quartz grain (shocked quartz) of a type formed only in large-meteorite impact craters, in the clay layer.

Before it settled to form the K–T boundary, the impact dust was transported all around the world by upper atmosphere winds until the whole Earth was enveloped in a dark cloud (Fig. 2.11c) which persisted for more than a year. Little solar radiation penetrated to the planet surface, the catastrophic fall in temperatures killed most terrestrial cold-blooded creatures and plants ceased to grow for lack of the light needed for photosynthesis. When the cloud finally began to thin (Fig. 2.11d) and the first pale rays of sunshine began to warm the planetary surface, the dinosaurs were dead. Some of the mammals, however, preserved from fatal chill by their homeothermic physiology and perhaps sustained by their opportunistic rodent-like dietary proclivities – grain and carrion were presumably available – survived the catastrophe. With their gigantic and frequently ferocious rivals now eliminated, the mammals seized their opportunity. They proliferated and evolved in all directions throughout the Tertiary period (65.5–1.81 mya) which had now commenced. The Age of the Mammals had arrived.

EVOLUTION OF HUMANKIND – THE ARBOREAL PHASE

When sunlight again flooded the land and the forests regrew, some small insectivorous mammals took to the trees to take advantage of the abundant insect life among the foliage, and also perhaps to avoid ground-dwelling predators. From these there developed during the Palaeocene epoch (65.5–55.8 mya), the first of the five epochs of the Tertiary period, a new order of mammals – the *primates* (Fig. 2.9d). As a result of the specific adaptive pressures of arboreal life, these became distinguished from other mammals in a number of ways. These included the presence of five digits

with free mobility (for grasping) on the extremities of their four limbs, some elaboration of the visual apparatus and, particularly significantly, a progressive elaboration of the brain – the cerebral hemisphere had large and prominent temporal lobes.

The first definite primate types were the *prosimians*, which are represented today by the lemurs of Madagascar, the lorises of tropical Africa and southern Asia, and the tarsiers of Borneo and other islands of the East Indies. They developed during the second epoch of the Tertiary period, the Eocene (55.8–33.9 mya), when flowering plants, which had originated as long ago as the Cretaceous period, began to flourish and became a major component of ecosystems previously dominated by the more ancient ferns, lycopods and conifers. The flowering plants greatly enriched the variety of food available in the forest canopy, not only by attracting more insects for pollination, but by supplying new edible forms of plant matter, especially fruit. By the time the next epoch of the Tertiary period, the Oligocene (33.9–23 mya) was under way, a new larger type of primate with wider dietary preferences, the *Anthropoidea*, had evolved from the prosimians. Their first representatives were the monkeys (Fig. 2.9d). Some time in the Eocene, perhaps within a diurnal (active during the day) species of prosimian[36] rather than the early nocturnal forms, there was a fortunate mutation which conferred an additional visual pigment in the retina. The retina became trichromatic rather than just dichromatic,[14] and it is from such a trichromatic individual that (Old World) monkeys are descended. The advantage of trichromatic vision is that it provides an improved ability to find ripe fruit and possibly also the young, more nutritious, leaves of tropical foliage, which are often red rather than green when they first emerge.[27] As we noted earlier, most mammal species are only dichromatic.

The species of the evolving monkey family occurred in a wide range of sizes as they do today. Since weight increases with the cube of linear dimension, but mechanical strength of a structure such as a tail increases only with the square, the larger monkeys became the less use their tail was in

climbing trees. In the fourth epoch of the Tertiary, the Miocene (23–5.3 mya) there eventually arose a large form without tail, which gave rise to what we now call the *tailless apes*, represented today by the gorilla, orang-utan, gibbon and chimpanzee. It was from an early tailless ape that humankind evolved.

EVOLUTION OF HUMANKIND – FROM APES TO MODERN MAN

On the basis of DNA comparison, and with plausible assumptions about rates of change of base sequence with time, it is thought that the divergence of those apes which were to give rise to humankind from those which were to give rise to our closest relative, the chimpanzee, took place about 7 mya. This, together with all the other major steps in the evolution of modern humans, took place in Africa. Those species which developed from this divergence point onwards, in the line of evolution to humankind, are lumped together with *Homo sapiens* in the taxonomic family the *Hominidae*, and are generically referred to as *hominids*.

Our knowledge of the intermediate steps in this evolution, especially the early ones, is fragmentary. As new fossils are unearthed, palaeo-anthropologists create new species and even new genera of pre-human hominids, and so this is a field which is highly fluid and becoming increasingly complex. The information yielded so far by the fossil record can be briefly summarized as follows. A skull of the earliest candidate progenitor after divergence from the apes was found in the Djurab desert of Chad in 2001 by a team led by French palaeontologist Michel Brunet. It is estimated to be 6–7 million years old, and has been given the name *Sahelanthropus tchadensis*. In Ethiopia, on the other side of Africa, fossils of a later hominid, *Ardipithecus ramidus*, have been found. The first to be discovered, by Tim White, appear to date from 4.4 mya, but what appears to be a much older example, from 5.2–5.8 mya, has been found by Yohannes Haile-Selassie. The best characterized example of a pre-human hominid, a nearly complete skeleton of which was found by Donald Johanson

and Tom Gray at Hadar in Ethiopia in 1974, has been given the name *Australopithecus afarensis*. This skeleton, possibly of a young female, is about 3.1 million years old. Other specimens of this genus have also turned up. In 1925 in South Africa, Raymond Dart found fossils of a species to which he gave the name *Australopithecus africanus* – the first use of the name *Australopithecus* ('southern ape'). These *A. africanus* fossils have been found in rocks 3–2.3 million years old, indicating that it came after *A. afarensis*.

By analogy with the 'mammal-like reptiles' discussed earlier, we can refer to these creatures as 'human-like apes' (Fig. 2.9e). Footprints preserved in volcanic ash from 3.6 mya, and the shape of the knee joint in *A. afarensis*, tell us that *Australopithecus* walked upright, a crucial distinction between this human-like ape and its 'ape-like ape' relatives and ancestors. *Australopithecus africanus* was about 150 cm tall and had a brain volume of ~440 cm^3, only about 30% of that of modern humans (~1450 cm^3) but comparable with the brain size of modern apes.

At least three *Australopithecus* species are known to have existed, and given the very poor fossil record it is likely that there were more. From one of these australopithecine species there arose, about 2 mya, a new hominid species with a larger brain (~640 cm^3), which had developed the ability to make simple stone tools. The first fossils were discovered by Louis and Mary Leakey at Olduvai Gorge in Tanzania in 1960, and shortly thereafter Louis Leakey, Philip Tobias and John Napier proposed that the fossils belonged to a new species, near enough to modern humans to be included in the same genus – *Homo*. The full name suggested for this perhaps ape-like human (as opposed to human-like ape) was *Homo habilis*, meaning 'able' or 'handy' human, to acknowledge its acquisition of tool-making ability. Hominids at the developmental level of *H. habilis* persisted in East Africa for about half a million years.

Not very long (in geological terms) after the arrival of *Homo habilis* there appeared a more advanced hominid, *Homo erectus* (Fig. 2.9e), the first member of our genus to master fire and also the first – perhaps with the help of fire to provide warmth and deter predators – to start living in caves.

H. erectus first appeared about 1.7 mya. This was the first hominid to spread beyond Africa – fossil remains have been found in Asia and possibly (although this is controversial) in Europe, the first known specimen being found in Java in 1891 by a Dutch army surgeon, Eugene Dubois. An almost complete skeleton, 1.6 million years old, of an adolescent male was found by Richard Leakey and Alan Walker[23] at Lake Turkana, Kenya, in 1984. *Homo erectus* was about the same height as modern humans. The brain size of *H. erectus,* at 800–1000 cm^3, was greater than that of *H. habilis* but still well below that of *H. sapiens.* Their tool-making was more advanced: hand axes with edges formed by removing stone from both sides were now being produced. It is plausible, too, that with the mastery of fire there came the discovery of cooking food, making it easier to chew and digest. This in turn made it less necessary to retain a strong jaw and big teeth for tearing and grinding food, and thus made possible the evolution of descendants with smaller teeth and less powerful jaws.

Those descendants appeared in due course. As well as having a reduced masticatory complex they came to have limb bones adapted to a fully erect posture and gait, a high forehead and a well-defined chin (compared to the sharply receding forehead and chin of *Homo erectus*), a greatly reduced brow ridge over the eye sockets, a more rounded and larger cranial vault and an average brain size of about 1350 cm^3, close to that of present humans. All these features finally came together somewhere in Africa 130 000–200 000 years ago, when evolution made the final adjustments to a new hominid species, *Homo sapiens,* that would become modern humans.

There must of course have been transitional forms between *Homo erectus* and *H. sapiens,* but the fossil record so far is rather reticent on this matter and indeed gives the impression that a fully functional *Homo sapiens* arrived all of a sudden as a complete package. One school of thought interprets the record as indicating the existence of (at least) one distinguishable species in the human line between *H. erectus* and *H. sapiens,* referred to as *Homo heidelbergensis.* Transitional stages between *H. heidelbergensis* and *H. sapiens* are thought to have existed ~250 000 years ago. Neanderthal

humans, lumbering, heavily built, beetle-browed, the quintessential Stone Age cavepeople beloved by cartoonists, are now thought not to be ancestral to humanity but to be a separate species, *Homo neanderthalensis*, diverging from the human line perhaps ~500 000 years ago and becoming extinct about 34 000 years ago.

Homo sapiens spread from its point of origin, perhaps East Africa, throughout the African continent and beyond, reaching the Middle East at least 90 000 years ago, Australia via Asia at least 46 000 years ago, Europe about 40 000 years ago, and the Americas (this is controversial) 13 000–40 000 years ago. The rest, so to speak, is history.

We have seen that the ape line of evolution forked 6–8 mya, one arm of the fork leading ultimately to humankind. What of the other arm, the subfamily *Paninae*, as opposed to *Homininae*? While the panin line no doubt gave rise to a variety of species, it is represented today only by two – the common chimpanzee, *Pan troglodytes*, and the pygmy chimpanzee, *Pan paniscus*. These are our nearest relatives in the animal kingdom. The first comparisons of DNA base sequence similarity indicated that we and the chimpanzees have about 98.4% of our DNA in common. This, however, is now looking like an overestimate. For example, Roy Britten has shown[3] that when insertions and deletions are taken into account, the divergence between chimpanzee and human DNA is about 5%. While accepting that there is much genetic similarity between the great apes and ourselves, I suggest that a great deal too much significance has been attached to this observation, ranging from the proposal that we should reclassify chimpanzees as belonging to the genus *Homo*, to the suggestion (in New Zealand) that chimpanzees and other apes should be given certain rights presently reserved to human beings.

There are two points to be made to put our relationship with chimpanzees into perspective. The first is that classification of living organisms is based always on the *phenotype* – on the observed morphology, chemistry, physiology and behaviour of the actual real-world living organism. Base sequences can be a powerful aid in tracing evolutionary pathways, estimating dates of evolutionary divergence and providing approximate numerical

values for relative relatedness of different species. But when it comes to biological classification, it is what we actually see, the nature of the organism that actually exists, that must take precedence, since it is not a base sequence but a living creature that we seek to place in our taxonomic system. It is on the sound taxonomic grounds of differences in structure, physiology, behaviour and mental capacity that chimpanzees and human beings have traditionally been placed in the separate genera, *Pan* and *Homo*, and the DNA base sequence gives us no reason to change this.

The second point is that in terms of its genetic implications the percentage difference between the DNA base sequence of the chimpanzees and ourselves is large, not small. There are about 30 000 genes in the human genome. Even if we choose the higher, although now superseded, estimate of genetic similarity a 1.6% difference could correspond to ~500 totally different genes, or it could correspond to very substantial differences in several times this many genes. Genetically speaking, a 1.6% difference in base sequence provides enormous scope for far-reaching differences between two species. The mistake has been to focus on the 98.4% (or 95%) DNA which is similar, instead of the highly significant and genetically very substantial 1.6% (or 5%) of the DNA which is different.

It should also be borne in mind that we share much of our DNA with many other organisms too, not surprising in view of the large proportion of metabolic and biosynthetic enzymic machinery that is the same in most living species. For example, the mouse – with which we had a common ancestor about 75 mya[10] – also has about 30 000 genes and ~99% of these have counterparts in the human genome. The average similarity between human and mouse genes is about 85%.[35] As an editorial in *New Scientist* (commenting on the New Zealand proposal) puts it:

> *A creature that shares 98.4% of its DNA with humans is not 98.4% human, any more than a fish that shares, say, 40% of its DNA with us is 40% human.*

EVOLUTION – DID IT REALLY HAPPEN AND, IF SO, HOW?

This story of the origin of humanity as presented by modern science, provisional and incomplete though it is in some parts, is not likely to change in its fundamental outlines, even though there is a vast amount of fascinating detail still to be filled in. This means that anyone who on the whole accepts the scientific world view really has no option but to accept that human beings evolved from apes, and beyond them from monkeys, prosimian primates, primitive insectivorous mammals, mammal-like reptiles, tetrapod amphibians, lobe-finned fishes, primitive fish-like vertebrates, early invertebrate chordates and so on back to the earliest forms of life 3.5 bya.

As we remarked earlier, it cannot be denied that this has been seen in the past, and still is to some people today, as an affront to human sensibility. The prospect that we, the pinnacle of creation, should have originated from such lowly ancestors is not always viewed with equanimity. Since the story of human evolution is always a very big pill, and sometimes a bitter pill, to swallow it is reasonable to demand evidence, both of its actual occurrence and of a mechanism by which the claimed evolutionary development could actually have taken place.

The evidence that biological evolution actually occurred comes from fossils. Fossils of early life forms really do exist, and the approximate times at which the corresponding living creatures lived can be determined with great confidence by examining the geological strata within which they are found, using all the techniques of sedimentary geology and geochemistry. Taking the evidence from detailed morphological study of these early life-forms, together with the temporal data, we are led inexorably to the conclusion that the living creatures which inhabit the world today, including *Homo sapiens*, evolved by progressive change from earlier, more primitive, forms. That biological evolution actually occurred is a brute fact. The evidence is all around us.

But if it occurred, how did it occur? What mechanism brought about the extraordinary proliferation of life-forms, and the continual replacement

of one life-form by another? The most commonly accepted explanation is Darwinian natural selection. This is by now so well known that we need say little about it here, beyond summarizing the essential principle. This states that in any given species, mutation of the genetic material (i.e. base sequence changes in the DNA of the chromosomes) caused by cosmic rays, solar ultraviolet radiation, errors in copying etc. give rise to minor variants in the progeny. Most of these are at a disadvantage, but occasionally one will occur which is better adapted to the surrounding environment; it will consequently multiply faster than its contemporaries. In this way, by the accumulation of small incremental changes, new species, new genera, new families and ultimately totally different life-forms supposedly arise from the old.

Does natural selection actually occur? Yes it does. The antibiotic-resistant bacteria that are now widespread in the developed world have arisen by preferential selection of resistant mutants during the course of repeated exposure of pathogenic bacteria to antibiotics since these medications were first introduced. The development of antibiotic-resistant strains may well also have arisen as a consequence of the irresponsible practice of including antibiotics in the diet of intensively reared livestock to make them grow faster.

The insecticide-resistant insects that now plague the cotton-growing industry have arisen by natural selection within the original, predominantly sensitive, populations as a consequence of their chronic exposure to insecticides over many years. The phenomenon of *industrial melanism* in the smoke-blackened industrial north of England may also be an example of natural selection – in a certain species of moth a dark-winged type, better camouflaged than its light-coloured relatives, has come to dominate the population within comparatively recent times.

Natural selection, it seems, can indeed happen. The problem is that all the observable changes within populations that we can confidently attribute to this process today are very small, frequently involving mutational change in just one gene. While accepting, as all biologists do, the

Darwinian insight that any genetic variants within a population which are better adapted to their environment than their peers will achieve greater reproductive success, are we nevertheless entitled to assume that the enormous changes in living organisms that have taken place, especially since the first multicellular forms arose ~700 mya, are due simply to the steady accumulation of randomly-occurring one-step-at-a-time mutations, *and nothing else*? Neo-Darwinian orthodoxy says that we are. Not all biologists, however, find this convincing. Eminent evolutionary biologist Ernst Mayr concludes from the fossil record that new species arise suddenly rather than by the accumulation of small changes. Similarly, Stephen Jay Gould and Niles Eldredge in their *punctuated equilibrium* version of Darwinian evolution argue that *speciation* (the process of creation of a new species) occurs as a relatively rapid event when certain organisms, for reasons that we can never retrospectively know, suddenly diverge from their parent population down a new genetic path. Indeed, we saw above that *Homo sapiens* seemed to arrive on the scene in a surprisingly sudden manner. Further questions have been raised by Manfred Eigen and Peter Schuster on the basis of mathematical modelling of natural selection within mutating populations. The results suggest that there are fundamental limits to the complexity of a genome (the complete set of genes possessed by an organism) that can be assembled by mutation and selection alone.

Given that we are dealing with unrepeatable historical events, it must be acknowledged straight away that we are unlikely ever to know with certainty whether the simple cumulative mechanism of the neo-Darwinians is in fact sufficient to create the enormous biological diversity that has arisen, or whether there must also be something else. Are there any plausible candidates for this 'something else'? One possibility is being explored by Stuart Kauffman[20] and colleagues at the Santa Fe Institute, and by Brian Goodwin[15] in England. They propose that within living organisms there are certain types of structure that, for fundamental physicochemical reasons, have a built-in or pre-programmed tendency to emerge as evolution proceeds.

A simple and uncontroversial example is the natural propensity of polar lipids – compounds having long hydrophobic hydrocarbon chains with a hydrophilic (water-attracting) chemical group at one end – to form, when suspended in water, thin 'bilayer' membranes with the hydrocarbon chains forming the inner regions of the membrane and the hydrophilic groups arrayed along each surface of the membrane where they can interact with the surrounding water molecules. These sheets of membrane can then, as we saw earlier (Fig. 2.3b), spontaneously curve back upon themselves and join so that some of the aqueous medium becomes enclosed within a membrane-bounded sac, or membrane vesicle as it is sometimes called. If we were able to travel 3.5 billion years back in time and examine ocean waters through a microscope we might well be able to see some of these tiny vesicles floating in the water. While we might reasonably hypothesize that these were early precursors of living cells, we would be quite wrong to say that they must have 'evolved' by natural selection. Their formation would have been a spontaneous process arising out of the physical chemistry of polar lipid molecules.

As Brian Goodwin puts it:

> *… morphogenesis is a process that has intrinsic properties of dynamic order so that particular forms are produced when the system is organized in particular ways.*

He discusses possible examples such as the principles underlying the fundamental similarity in limb structure in different tetrapod animals, and the idiosyncratic and apparently adaptationally irrelevant morphology of the alga, *Acetabularia.* Goodwin looks forward to a day when:

> *Biology would begin to look a little more like physics in having a theory of organisms as dynamically robust entities that are natural kinds, not simply historical accidents that survived for a period of time.*

Or as Stuart Kauffman puts it:

> *Self-organization may be the precondition of evolvability itself. Only those systems that are able to organize themselves spontaneously may be able to evolve further.*

This view of evolution suggests the fascinating possibility that there exists, in the same way that mathematical truths exist before we discover them, an array of 'virtual blueprints' for possible organisms. That when random mutation allows a group of organisms to stray within a certain genetic, biochemical and morphological distance of one such blueprint the pieces fall into place rather quickly, a novel but inherently probable structure emerges, and a new species is born. In the terminology of dynamical systems, each virtual blueprint acts as an attractor in state space. Such a fundamental mechanism might underlie the phenomenon of *evolutionary convergence* which, quoting Cambridge palaeobiologist Simon Conway Morris in his book *Life's Solution: Inevitable Humans in a Lonely Universe*, is 'the recurrent tendency of biological organization to arrive at the same "solution" to a particular "need"'.

He gives numerous examples of where a certain structure or function – e.g. sight, echolocation, viviparity (live births) etc. – have arisen time and time again in the history of the biosphere by quite different evolutionary pathways. A particularly striking case is that of the now sadly extinct *Thylacinus cynocephalus*, the 'Tasmanian tiger' (Fig. 6.3). It looked and behaved like a wolf and occupied a similar ecological niche, but was in fact a marsupial not a placental mammal. To find the last common ancestor of wolves and thylacines we have to go back about 140 million years.[10] A fascinating but unanswerable question is whether, if we could go back in time, say to the first prokaryotes, and let evolution proceed again, we would eventually arrive at a broadly similar set of life-forms or something totally different. Stephen Jay Gould argued that every rerun would give a totally different result. Simon Conway Morris, by contrast, believes that the underlying tendencies for convergence are so strong that something approximately similar would eventuate each time. Perhaps natural selection, operating upon the enormous variety of forms generated by random mutation, is the mechanism by which life finds its way to the set of preordained niches that were all implicit in the Big Bang.

Whatever the actual mechanisms, whether or not we will have to invoke something additional to simple neo-Darwinian mutation and selection, it is universally agreed within biological science that the living organisms of the present day, and the extinct forms in the past, did indeed evolve from pre-existing forms by processes explicable in terms of the laws of physics and chemistry. Some commentators jump immediately from this to the view that humankind is therefore a mere biochemical accident in a meaningless universe; that humanity with all its beliefs, emotions and culture is just one more ephemeral collection of atoms of no fundamental significance. This is simply one metaphysical view of the nature of reality which, contrary to what is sometimes assumed, is given no more support by science than a variety of other metaphysical views. We return to this topic in the last chapter.

Chapter 3

The Nature of Physical Reality

… and gives to airy nothing a local habitation and a name.

William Shakespeare, *A Midsummer Night's Dream*

Anyone who is not shocked by quantum theory has not understood it.

Niels Bohr (a founder of quantum physics)

Of all the 'Big Questions' the most fundamental is – what is the nature of reality itself? This has been a constant preoccupation, since earliest times, of philosophers in the West and of Vedic and Buddhist thinkers in the East. In the account given so far of the origin of the physical universe as we see it around us on any starry night, and of the origin of our sun, our planet and its biosphere of which we form a part, there has been the implicit assumption that there really is a physical/chemical/biological universe out there, with clearcut well-defined properties to which our sense impressions give us access, and about which we can therefore make confident assertions. Although this is the way most of us talk most of the time, the picture of physical reality, which science at its most fundamental level has revealed over the last 100 years, does not support this simple everyday view. Certainty has been replaced with unavoidable, we might even say *compulsory*, uncertainty. As we shall see in this chapter, the closer we look at the physical world it becomes not more clearly defined but more nebulous, seeming almost to disappear from view.

Thinking people have known for centuries that physical reality cannot be identical with our perception of it. The 17th century English philosopher John Locke (1632–1704), for example, argued that material objects do not in themselves possess colour, sound, taste or smell – these are what he called *secondary* qualities that only come into being in the conscious mind of the human observer. As Galileo (1564–1642), a founder of modern science, put it, concepts such as colour 'hold their residence solely in the sensitive body', i.e. the observer. Nevertheless, assuming we adhere, as most people and certainly most scientists do, to a realist philosophy (the belief that physical objects exist independently of being perceived by an observer), we may plausibly assume that material objects have their own *primary* properties, and these, by mechanisms to be ascertained, give rise to the sensations of secondary properties in the observing human.

The study of these hypothesized primary properties of the material world is the study of the nature of physical reality: this is the business of science. Taking pre-Socratic Greek philosophy as the starting-point of rational consideration of these matters, what does two and a half millennia of science have to tell us about the nature of physical reality?

THE WORLD VIEW OF CLASSICAL PHYSICS

Around the end of the 19th century, when what is called *classical* physics still prevailed, we would have got a clear answer to this question. The physical universe was made up of matter and radiation. The matter was composed of various kinds of atoms, which could be linked in different ways to form molecules. Material bodies attracted each other in accordance with the law of gravitation. The radiation consisted of an electromagnetic disturbance, of a wavelike character, propagating through space. Matter and radiation both behaved entirely in accordance with the deterministic laws elucidated by scientists such as Galileo, Newton, Faraday, Maxwell, Coulomb, Gauss and many others. In principle, as most clearly enunciated by French math-

ematical physicist Pierre Laplace (1749–1827), the future state of the universe at any time was entirely determined by its state at the present, and it would inexorably evolve towards that future state in accordance with the wholly deterministic laws of science as then understood.

At the turn of the 20th century there were still many unresolved problems: scientists of the time were well aware that science was not complete. But they were confident that the remaining problems would eventually be solved, and moreover that the solutions would be within the framework of classical physics.

What happened, however, was that in the first few decades of the 20th century many of the major problems of physics were indeed solved, but at great cost – nothing less than the abandonment of the deterministic world view that had held sway in physics throughout its classical period. I am not referring here to relativity theory which, despite the seemingly bizarre notions of time passing at different rates depending on how fast you are travelling, or of gravity really being due to a curvature of space-time, is properly seen as being part – although a very novel and revolutionary part – of classical physics. Determinism rules in both the special and the general theories of relativity. What caused the foundations of classical physics to crumble was a new physical theory, which came to be known as *quantum mechanics*.

The discovery of quantum mechanics

This theory initially arose out of attempts to explain the observed spectral energy distributions (the way energy output varies with wavelength) of bodies such as the sun, or indeed any hot object, which emit radiation. The shapes of these spectral distributions predicted by physical theory as it existed at the end of the 19th century were quite different from what was actually observed. The problem was solved in 1900 by German physicist Max Planck (1858–1947) by the novel but seemingly innocuous proposal that radiant energy was emitted and absorbed – not in indefinitely large or

small amounts at a time, as had been assumed – but only in discrete units, or *quanta*, of energy the size of which was a simple function of wavelength in accordance with:

$$\varepsilon = \frac{hc}{\lambda}$$

where ε is the energy in one quantum of radiation, λ is the wavelength of light or other radiation, c is the speed of light and h is a new physical constant, now known as Planck's constant. Planck was able to show that once this assumption was made, an equation for the spectral distribution of radiant bodies could be arrived at whose predictions accorded very well with observation (Fig. 3.1).

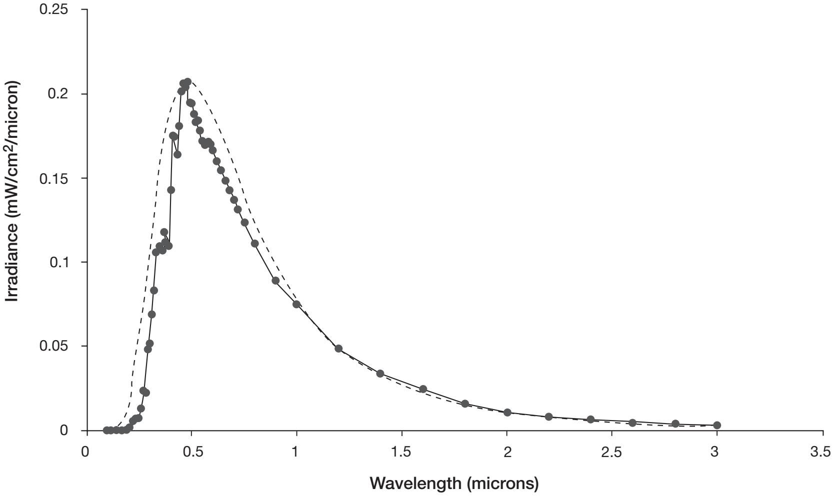


Fig. 3.1 Spectral energy distribution (•) of solar radiation (in mW cm^{-2} μm^{-2}) measured above the Earth's atmosphere, together with that calculated for an idealized black body at 6000 K using the Planck equation. Before quantum theory, classical physics predicted a distribution continuing to rise indefinitely with decreasing wavelength; instead, there was an observed peak. The solar spectrum differs somewhat from that of an ideal black body because of the absorption bands of hydrogen in the Sun's outer atmosphere.
Source: Data from Thekaekara, M.P. & Drummond, A.J. (1971). *Nature* **226**, 6–9.

Planck's new theory was hailed as a major breakthrough, as indeed it was; it was also, although no-one realized it at the time, the first step down a slippery slope at the bottom of which lay a totally new, strange and counterintuitive view of the nature of the world. The evolution of what came to be known as quantum mechanics is a fascinating story, and undoubtedly one of the most important in the history of science. To do it justice, however, would take a great deal of space, and accounts have already been written by others,[2,12] so in this chapter we will proceed straight to a consideration of the implications of quantum mechanics, in the form in which it exists at the beginning of its second century, for the fundamental nature of physical reality. On the historical side, suffice it to say that as the pioneer quantum physicists pursued the implications of the quantization of energy, especially when combined with the fact that radiant energy also has unmistakable wave-like properties, they were drawn further and further into a strange world of a nature quite alien to that of the classical physics in which they had all been trained.

The first thing to make clear about quantum mechanics is that it leaves the classical physics description of most aspects of the macroscopic everyday world quite untouched. The laws hypothesized to exist, and the equations used to calculate behaviour, for anything we can see or feel, or indeed even see down a microscope, are still mostly those of classical physics. The realm where quantum mechanics holds sway is the realm of the ultra-small: the individual atom, molecule, subatomic particles (electrons, protons etc.) or photons (a quantum of radiation). The fact that it deals primarily with things that cannot be part of our everyday experience does not mean that we can ignore quantum mechanics. Not only can events on the quantum scale sometimes be amplified up to the macro scale, but certain important and useful macroscopic phenomena (transistors, lasers, superconductivity, nuclear fusion etc.), not to mention the spectral energy distribution of the Sun, can only be understood in quantum mechanical terms.

QUANTUM MECHANICS AND DETERMINISM

So what does quantum mechanics have to say about what is happening in the ultra-microscopic world? The most extraordinary claim that quantum mechanics makes is that individual quantum events cannot be ascribed to a particular cause. Consider, for example, the phenomenon of radioactive decay. The atoms of certain elements are unstable: they have a tendency to disintegrate into smaller atoms, typically with the simultaneous liberation of highly energetic subatomic particles – alpha particles (naked helium nuclei), electrons, positrons – and electromagnetic radiation in the form of gamma-rays. A well-known example is natural uranium, which contains very small amounts of two unstable isotopes, uranium-234 (0.0057%) and uranium-235 (0.72%), mixed with the stable isotope uranium-238. The particles and gamma-rays emitted during radioactive decay have such high energies that the disintegration of a single atom can be detected by appropriate electronic means, such as a Geiger counter.

If you were to place the probe of a Geiger counter over a very tiny piece of uranium you would, if the instrument had an audio output, hear a series of clicks corresponding to the disintegration of U^{234} and U^{235} atoms. As you continued listening you would notice that the clicks followed no pattern in time. They would not occur at regular intervals but would appear to be quite randomly spaced. And if you took the trouble to actually measure the intervals between clicks and subject the data to statistical analysis, you would find that the clicks, and therefore the atomic disintegrations, were indeed occurring in an entirely random manner with respect to time.

These observations, however, should not be taken to imply that the phenomenon of radioactive decay is entirely devoid of regularity or law-like behaviour. On the contrary, if you were to set the instrument to count the large number of disintegrations that took place over some substantial period, say one hour, and you repeated this experiment many times, you would find that there was very good agreement between the different hourly counts. While it is unlikely (although possible) that any two counts

would give exactly the same number, when expressed as a proportion or percentage they would differ little from each other.

Furthermore, if you were to repeat the whole experiment with a similar amount of a different radioactive element, such as radium, you would again find that despite randomness at the level of individual disintegrations, there would be good agreement between successive hourly totals. You would further find that after making appropriate allowance for the non-participation of the stable U^{238} isotope, the calculated hourly rate of disintegration per microgram of radioactive element exhibited by the U^{234}/U^{235} mixture was quite different from that of radium. That is, each radioactive element has its own characteristic, and entirely reproducible, rate of decay. Such rates are commonly expressed in terms of the *half-life*, which is the time it would require for half of any given sample to decay; this is quite independent of the size of that sample. For the radioactive elements we have been discussing the half-lives are uranium-235, 713 million years; uranium-234, 248 000 years; radium, 1622 years.

How do we reconcile these reproducible quantitative differences between radioactive elements with the entirely random occurrence in time of the individual atomic disintegrations? According to quantum mechanics, no individual atomic disintegration can meaningfully be attributed to any specific cause – hence the entirely random distribution in time. However, the atoms of any given radioactive element will have a certain built-in propensity to disintegrate which manifests as a rather constant rate of decay characteristic of that element when measured over a sufficient period. An integral part of the decay process, very often, is the emission or escape of an alpha particle from the nucleus. Quantum mechanics represents all particles as a form of wave energy (discussed later), and the nature of wave energy is to be, to a greater or lesser extent, spread out in space, unlike particles which we usually think of as having precise locations. For any particle in a given situation there is an equation – the *Schrödinger wave equation* (see Box, p. 113) – which describes the distribution in space of a particular attribute of the particle, known as the *wavefunction* and indicated by the

Greek letter ψ (psi). The true physical meaning of ψ was a highly controversial matter in the early days of quantum mechanics, but the prevailing interpretation is that the square of the value of the wavefunction (strictly speaking, the square of the modulus of the amplitude of the wavefunction) at a particular location in space is proportional to the *probability* of the particle actually being at that location in space. More rigorously, we should refer to the probability of the particle being *found* at that location, in the event of a measurement of its position actually being carried out.

Normally an alpha particle is firmly bound within the nucleus, held by a force known as the *strong nuclear force*. But calculation of the wavefunction of an alpha particle in an unstable nucleus of the kind we have been discussing shows that it has a tiny, but non-zero, value outside the barrier presented by the strong nuclear force. This means that there is a corresponding very small probability of the alpha particle being located outside the barrier, and every now and then an alpha particle will realize this probability and escape from the nucleus. This is a process referred to as *quantum tunnelling*, because it is as though the particle tunnels through the barrier (Fig. 3.2). In quantum mechanical terms, then, the differing stabilities of the three radioactive elements discussed above (the built-in differences in their propensity to decay), is an expression of differences between the wavefunctions of alpha particles in the respective nuclei.

From the point of view of the macroscopic world that we inhabit there are two things about the quantum mechanical description of radioactive decay, and other processes at the ultramicroscopic level, which are very strange. The first is that, in a manner decreed by the actual shape of the wavefunction, i.e. the variation of ψ with position in space, a given particle at a given point in time can find itself located just about anywhere, even on the other side of a barrier which might otherwise be regarded as impenetrable. It is as though I were to throw a tennis ball against a brick wall, and suddenly find that it had actually gone through to the other side. We may wonder – if it is true for alpha particles in nuclei, why should it not also be true for tennis balls in the backyard? The answer is that it is true for tennis

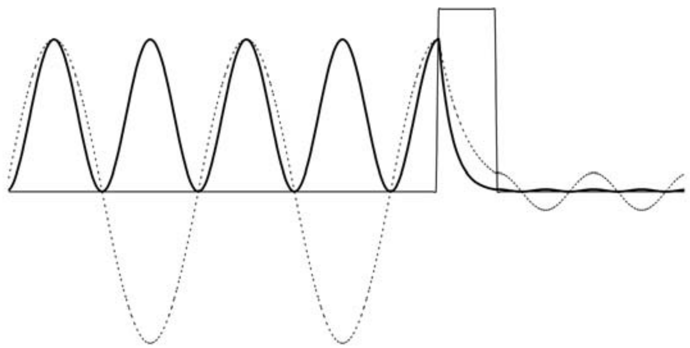


Fig. 3.2 Quantum tunnelling. A particle traveling from the left with total energy E, encounters a barrier which would require energy V to penetrate where $V > E$. Here the particle is shown as a wave packet, and the broken line (-----) is $\psi(x)$, the value of the corresponding wavefunction as a function of distance, x, in accordance with the Schrödinger wave equation, at a given point in time. Within the barrier the wavefunction does not fall to zero but diminishes exponentially, reaching a very low but finite value at the outer edge of the barrier. From there it once again takes on the normal form of the Schrödinger wave equation for a freely travelling particle. The thick line is $|\psi(x)|^2$, the probability of the particle being found at any given value of x.
Source: Moore, W.J. (1956). *Physical Chemistry*. Longman: London, pp. 277–280. See also Davies, P.C.W. (1984). *Quantum Mechanics*. Routledge & Kegan Paul: London, pp. 42–44.

balls in backyards. However, when the quantum mechanical calculations are carried out for macroscopic objects such as tennis balls (which are about 10^{25} times as heavy as an alpha particle) the wavefunction is such that the probability of passing through a brick wall is so low (although never zero) that it is unlikely to have happened even once in the whole history of the universe.

The second strange thing is the assumption that individual events, such as the escape of an alpha particle from a nucleus, or the jump of an electron within an atom from one orbit to another accompanied by the emission of a light photon, are entirely probabilistic in nature. That while a certain characteristic average frequency of these events will be observed over time in accordance with the appropriate wavefunctions, no individual event can be said to have any particular cause. We are of course very familiar

with probabilistic processes, where the laws of probability apply in the macroscopic world: the tossing of a coin or the rolling of dice are the classic examples. But even though the outcomes of coin-tossing or dice-rolling trials are routinely found to be such that apparently only chance prevails, classical physics would nevertheless argue that every individual event proceeds precisely in accordance with the deterministic laws of Newtonian mechanics. The apparently random outcomes are attributed to the large number of unknown, uncontrolled and independently varying physical parameters of the system: the precise upward force applied to the coin by the thumb, the exact position on the coin where the force is applied, the position on the curved index finger where the coin rests before tossing, the height of the hand above the floor, the viscosity of the air as determined by temperature and humidity, localized air currents, turbulent aerodynamic processes around the surface of the spinning coin, friction between the coin edge and the floor on impact, locally varying elasticity of the floor and so on. In short, the apparently random behaviour of the tossed coin or rolled dice is attributed to the independent operations of a set of variables of the system.

Might not the same be true of events at the quantum level? Might it not be the case, for example, that this particular U^{235} atom on the left disintegrated a moment ago, and not the one on the right, because the values of certain presently unknown physical variables affecting the behaviour of alpha particles within the nucleus were different in the two nuclei, in such a way that escape of an alpha particle was momentarily possible in the left-hand nucleus but not in the right-hand one? This is an example of *hidden-variable theories*: these are theories whose aim is to restore determinism to quantum physics by postulating that the apparently random behaviour of quantum systems is in fact due to the operation of some inner variables of the system, of which we are as yet ignorant (hence the name *hidden* variables). Hidden-variable theorists would assert that physical systems which we think are in an identical state, such as two U^{235} atoms side-by-side, or two molecules both with an electron excited to a particular orbit, may not be

identical because they have different values of the hidden variables; this is the explanation of why just one of the U^{235} atoms disintegrates and just one of the excited molecules emits a photon within a specified time period.

Let us acknowledge straight away that hidden-variable theories are entirely permissible within physics. Determinism has been so central a part of classical scientific thinking that many scientists were very unhappy about abandoning it, and even the rather ad hoc and arbitrary nature of hidden-variable theory has seemed to some an acceptable price to pay in order to retain the deterministic world view. Nevertheless it must be said that the most widely accepted position within theoretical physics is that individual events at the quantum level are inherently undetermined, that their apparent randomness within the statistical constraints imposed by the wavefunction reflects an unpredictability which is fundamental, not merely a consequence of our ignorance of the operation of controlling variables hidden from our gaze within the system.

For events down at the level of the fundamental particles of which the universe is composed it seems, if the prevailing version of quantum mechanics is to be accepted, that we are called on to abandon the familiar notions of causality that all of us hold on the basis of the way we observe things in the everyday macroscopic world. But that is not all that quantum mechanics requires us to abandon. It rejects the very idea that at a given instant in time a physical entity such as an electron or alpha particle can be said to actually have a specific exact location in space, and to be travelling with a specified velocity.

HEISENBERG'S UNCERTAINTY PRINCIPLE

We touched upon this aspect of the theory earlier when discussing the wavefunction, the value of which for a particular particle at any point in space is related to the probability of the particle being found – supposing we set up an experiment to actually locate it – at that point in space. The

idea that no meaningful attribution of an exact simultaneous position and velocity – or more usually position and *momentum* (which is simply velocity × mass) – for a particle can ever be made first arose in a seemingly uncontroversial manner from the realization by Werner Heisenberg (1901–1976), one of the founders of quantum theory, that particles at the atomic level are so small that any attempt to measure their position and velocity, using photons of light or other electromagnetic radiation, would disturb the observed particle to such an extent that it would be impossible to know for certain just where it had been at the moment of observation. In 1927 Heisenberg analysed this problem using the theory of quantum mechanics that had been developed thus far, and concluded that the unavoidable uncertainty in the measurement of position and velocity (or momentum) could be expressed quantitatively by a simple equation:

$$\Delta p \Delta q \geq \frac{h}{2\pi}$$

which in effect says that the uncertainty in momentum (Δp) and the uncertainty in position (Δq) are always such that their product must be at least as great as the quantity h (Planck's constant) divided by 2π. He arrived at a similar relationship between the uncertainties in the energy (ΔE) of a particle, and the time (Δt):

$$\Delta E \Delta t \geq \frac{h}{2\pi}$$

These relationships are commonly referred to as the *Heisenberg uncertainty principle.*

On the face of it, while somewhat discouraging in so far as it puts a limitation on what we can measure, the principle does not have any immediately obvious philosophical implications. It could be seen as expressing in mathematical form the quantitative character of a purely practical problem. A similar problem was already recognized in the field of light microscopy, where it was well known that with light of a certain wavelength

objects below a certain minimum size cannot be resolved no matter how powerful the lenses of the microscope. For Heisenberg, however, and the other founders of what is now the prevailing school of quantum mechanics, especially Niels Bohr (1885–1962), the significance of the uncertainty principle is that the particle simply does not have a definite position or velocity. When we choose to measure one or the other, what we are doing is forcing the particle to abandon its inherently indeterminate nature and occupy just one position, or exhibit just one velocity, within the range of values permitted by the uncertainty principle. As soon as we measure one quantity exactly – e.g. position, so that $\Delta q = 0$ (no uncertainty in position) – then the velocity or momentum becomes completely uncertain (Δp is infinite), and vice versa.

Waves and Particles

Bohr's preferred version of this interpretation was to link uncertainty to the fact that objects at the quantum level can behave either as waves or particles, depending on the methods by which we observe them. A beam of light shining on a photoelectric cell, for example, seems to be particulate in nature – photoelectricity can only plausibly be explained by the hypothesis that the light consists of a stream of photons and each photon causes the ejection of one electron from the metal surface. If on the other hand a beam of light is shone on two narrow parallel slits in a screen, then on another screen placed beyond the first one, a so-called interference pattern (alternate light and dark bands) is formed (Fig. 3.3), and this can only be explained by the supposition that light is a form of wave energy. How something can simultaneously be both a particle and a wave remains one of the central mysteries of quantum mechanics, but the experimental evidence is so strong that we must accept that it is so. Particle-wave duality has been observed not only for photons, but also for electrons, neutrons, atoms and diatomic molecules. More remarkably still, it has now been

demonstrated for the very substantial molecules of *fullerene*, which contain 60 or 70 carbon atoms. Anton Zeilinger and co-workers showed that a collimated beam of fullerene molecules passed through a diffraction grating with 50 nm-wide slits developed an interference pattern on a detector placed behind the grating.[1]

The significance for the uncertainty principle of wave-particle duality is that while a particle can readily be imagined as having a precise location, wave energy can be spread out in space, so to the extent, for example, that we force a quantum object to occupy a definite position (by carrying out a position measurement) we may be eliciting its particulate rather than its wave aspect. As an in-between situation we can envisage a quantum object as a *wave packet* – a short train of waves extending over a limited region in space.

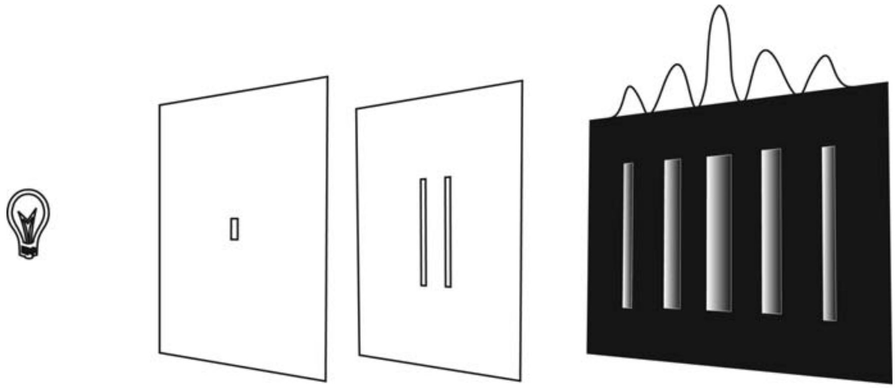


Fig. 3.3 Double-slit interference phenomenon due to the wave nature of light. Light from a point source illuminates a screen with two very narrow parallel slits close together. On a second screen (behind the first) there appears a central bright line with a series of parallel alternating dark and bright lines on either side. These are referred to as 'interference fringes'. At any given point on the screen the light waves arriving from slit 1 and slit 2 have traversed slightly different pathlengths so that they can arrive in phase (thus adding together to form bright fringes – constructive interference) or out of phase (a peak in one wave coinciding with a trough in the other), thus cancelling out to give dark fringes (destructive interference). The curve on top of the screen is a graph of the distribution of light intensity.

The inherent non-localizability of objects of atomic size decreed by quantum mechanical theory in fact extends well beyond that implied by the Heisenberg uncertainty principle. We recall that the wavefunction of a particle has a form determined by the Schrödinger wave equation which specifies the distribution in space of the wavefunction, ψ: this determines the probability of finding the particle, if such a measurement is carried out, at any given location in space. We saw that this can be used to explain the escape of an alpha particle from the nucleus of an unstable atom, but in this case we are dealing in uncertainties of position only of atomic dimensions. Consider another physical system, namely a beam of light incident on a half-silvered mirror, i.e. a mirror which reflects half the light and lets the rest pass through (Fig. 3.4). Let us suppose that we also have suitably placed detectors with which to measure the reflected and the transmitted light. If we consider the light energy to be entirely particulate in nature, we will suppose that half the particles pass through the mirror and half are reflected: any individual particle, i.e. photon, will either be reflected or transmitted. However, according to quantum mechanics the correct wavefunction of a single photon after interaction with the mirror is in fact the sum of two wavefunctions – that for a transmitted photon and that for a reflected photon. This is known as the *principle of superposition.* This means that the photon is in both the transmitted and the reflected beams of light. However, when we make a measurement with the detectors, then all the energy of the photon will be found to be in just one of the two beams. Direct experimental evidence that single photons can be delocalized in this way is provided by the double-slit light interference experiment referred to earlier. Even when the light intensity is reduced to the point that only one photon can pass through the slits at a time, an interference pattern (which requires light to come from both slits) still eventually builds up on a photographic plate placed beyond the screen. Each photon is, so to speak, interfering with itself. Similar experiments have been carried out with a beam of electrons – quantum theory says that even classically particulate objects such as electrons, protons and neutrons etc.

are waves as well as particles. Once again, even when the electrons pass through the double-slit apparatus (in this case, a device called an electron biprism) one at a time, an interference pattern builds up on the detection screen (Fig. 3.5). Thus, unless we invoke some additional agency acting on the system to make it behave in this way, we are led to the conclusion that any given photon or electron must somehow pass through both slits. But when each photon or electron finally arrives at the detecting screen, it acts unequivocally as a particle, making its impact just at one point.

When it was first proposed, the Heisenberg uncertainty principle might reasonably have been seen simply as an expression of the practical limita-

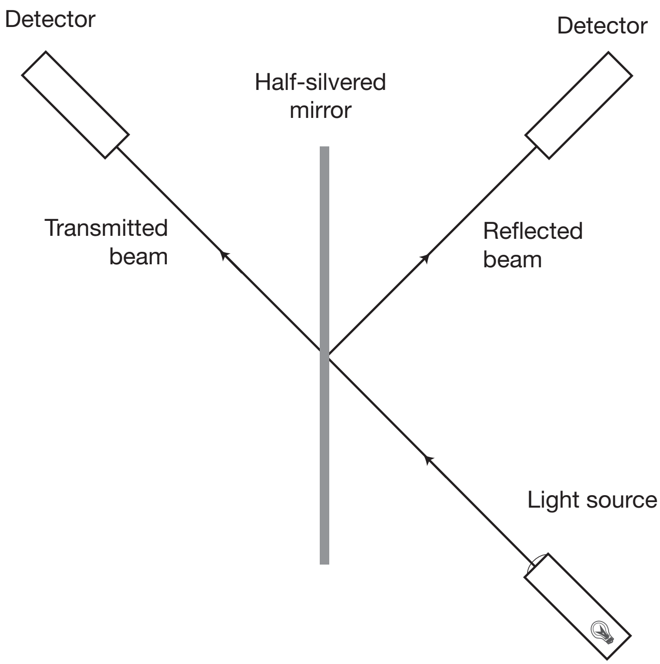


Fig. 3.4 Principle of superposition. A parallel light beam is incident at 45° on a half-silvered mirror, i.e. a mirror which reflects 50% of the light and transmits the rest. Photon detectors can be positioned so that they measure the transmitted and the reflected light. The correct wavefunction of any single photon is the sum of half the wavefunction of a transmitted photon and that of a reflected photon. However, if measurements are actually made then any given photon is shown to be either transmitted or reflected, not a mixture of both.

tions on our accuracy of measurement. In modern quantum mechanical theory, however, it is interpreted, not merely as a limitation, but as a creative principle. On the basis, for example, of the $\Delta E \Delta t$ uncertainty it is believed that at the quantum level the law of energy conservation can be suspended for very short periods of time. As theoretical physicist Paul Davies puts it:[8]

> *During this brief duration energy can be 'borrowed' for all manner of purposes, one of which is to create particles. Any particles produced in this way will be pretty short-lived, because the energy tied up in them has to be repaid after a minute fraction of a second. Nevertheless, particles are permitted to pop out of nowhere, enjoying a fleeting existence, before fading once again into oblivion.*

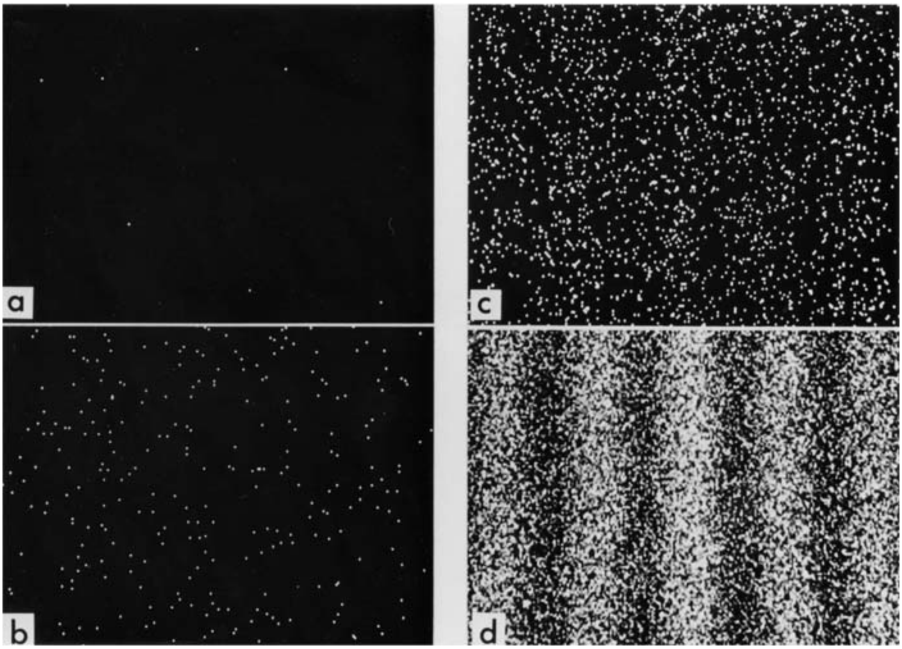


Fig. 3.5 Wave-particle duality. After passing through a double slit, electrons arrive on the screen one at a time as individual particles. As their numbers build up, the banded interference pattern, characteristic of wave energy, emerges.
Source: Reprinted courtesy of Dr Akira Tonomura, Hitachi Ltd, Japan.

This is the mechanism on the basis of which the quantum vacuum (which we discussed in the first chapter) is believed to be a 'seething ferment of virtual particles'.

NON-LOCALITY

As well as telling us that any atomic-scale objects must be regarded as being in some sense simultaneously present throughout a region of space much larger than the object itself turns out to be when we actually observe it, quantum mechanics also claims that certain categories of events can involve causally linked changes in pairs of particles, which occur simultaneously even though the particles are separated by distances that can be indefinitely large, e.g. of interplanetary dimensions. A comprehensive explanation would take a great deal of space, so I give here a highly abbreviated account.

Particles at the quantum level have a property that is referred to as spin: this is not identical to spin as we know it in everyday macroscopic objects, but has some similarities. Quantum spin has numerical values which are always integer or half-integer multiples of the quantity h (Planck's constant) divided by 2π, the permitted values of spin being 0, $1/2(h/2\pi)$, $1(h/2\pi)$, $3/2(h/2\pi)$, $2(h/2\pi)$. Any given type of particle only has a certain value of spin, e.g. electrons have spin ½. In the case of macroscopic objects spin has direction: this can be thought of as the direction the axis of rotation is pointing, up or down, indicated by + or -. When a measurement of spin along a specified line chosen by the experimenter is carried out on a quantum object it will be found to have just one of the discrete set of permitted values referred to above, but the direction of spin will point either forward or backward along the line. If the particle is an electron, for example, its measured spin will be found to be either +½ or –½, although for any given particle there is no way of predicting which result will be obtained.

We consider now the situation in which an unstable particle in the laboratory decays into two other particles, identical to each other, which fly off in opposite directions. These particles both have spin, but the original unstable particle has zero spin. After the particles have travelled some distance we measure the spin of one of them, say the left-hand particle, along a specified direction. Before the measurement, whether the spin will turn out to be +½ or –½ is genuinely undetermined, although with differing probabilities that can be calculated in advance. The same is true of the right-hand particle. However, quantum mechanics tells us that as soon as we measure the spin of the left-hand particle, then the spin of the right-hand particle measured along the same direction, no matter how far away it is by now, instantly ceases to be undetermined and takes on a definite value – it is numerically equal but opposite in sign to that of the left-hand particle. It has to be of opposite sign because the original unstable particle had spin zero. Angular momentum must be conserved so if one daughter particle has spin +½, the other must have spin –½ to cancel out.

Before measurement, the left-hand particle was neither entirely spin +½ or spin –½: its state was describable by a wavefunction which contained both possibilities. By making a measurement we force the particle to occupy one or other of its two options. When we do so we simultaneously force the right-hand particle to choose one of its options, the one of opposite sign to that arrived at by the left-hand particle.

If we do not want to accept quantum mechanical uncertainty as the explanation for this kind of phenomenon, and try to apply the hidden-variable approach, we might argue that the two particles were already in the spin +½ and spin –½ states before measurement. When we measure one to be spin +½, then we know straight away that the other must be spin –½. However, a remarkable theorem arrived at by Northern Ireland theoretical physicist John Bell in 1964 showed that quantum mechanics and any such hidden-variable theory would give rise to different statistical distributions of results if measurements of this type were actually carried out. The matter was put to the test in various laboratories, most successfully by Alain

Aspect and his colleagues in Paris in 1981. They used the polarization states of pairs of photons rather than, say, the spin of pairs of electrons, but the principles are exactly the same. The results unequivocally supported the quantum mechanical rather than the hidden-variable view of what happens when measurement is carried out on a pair of particles whose properties are linked in the manner discussed. Such pairs of quantum objects – particles, atoms or even larger systems – are referred to as being in an *entangled state,* and in quantum mechanics are treated as single, rather than separate independent objects.

The general phenomenon of apparently instantaneous changes occurring over significant distances when a measurement or other interaction takes place in just one part of an entangled system is the phenomenon of *quantum non-locality*. Interestingly, this and certain other aspects of quantum mechanics such as indeterminism were anathema to Albert Einstein even though his early work in photoelectricity played a key role in re-establishing the view that light energy consists of freely travelling individual quanta, or photons – one of the cornerstones of quantum theory. It is particularly ironic that experiments of the Aspect type, which vindicated non-locality, can be traced back conceptually to a thought experiment concocted by Einstein with his colleagues Podolsky and Rosen, specifically with the aim of persuading people that since quantum mechanics implies non-locality – which at first sight seems ridiculous – then the theory must be grievously at fault.

The Measurement Problem

We talked earlier in this chapter about measurements on quantum systems, and that when we measure the position of a particle we force it to occupy just one position out of the range available to it. In the half-silvered mirror experiment, when we carry out a measurement with the detectors, we force a photon to be in either the reflected or the transmitted beam, not both. What is going on here?

According to quantum mechanics, before the measurement is made, e.g. in the case of the half-silvered mirror, the photon exists as a superposition of two quantum states, one of which corresponds to the photon having been reflected and the other to it having been transmitted. The wavefunction, ψ, of the photon is a combination of the separate wavefunctions of a reflected and a transmitted photon. The two separate wavefunctions, and their combination, evolve in time in accordance with the Schrödinger equation. When, however, we perform a measurement with the detectors we do not find part of the photon's energy registering on each, as the wavefunction description might lead us to suppose. Instead, the photon is found to have delivered all its energy to one detector or the other, as though the photon was entirely reflected or entirely transmitted. It seems that when a measurement is carried out, the wavefunction 'collapses' from its spread-out superposed state, to just one localized state which is registered on the macroscopic measuring apparatus.

This *collapse of the wavefunction*, or *reduction of the state vector* as it is sometimes called, is something that happens when a measurement of an event or situation at the atomic, or quantum, level is carried out with an apparatus which functions at the macroscopic, or classical, level. It is as though as soon as a quantum object finds that we are looking at it, it abandons its quantum mechanical lifestyle and starts behaving like a classical object. Quantum mechanics itself provides no explanation of why this should happen: the equations do not predict it. Indeed, so far as the theory is concerned the superposed wavefunction for a reflected and a transmitted photon just keeps on evolving in the manner prescribed by the Schrödinger equation. In its standard form, quantum mechanics provides no mechanism by which the wavefunction can be made to collapse.

This collapse of the wavefunction when a quantum state is observed with a measuring apparatus at the macroscopic level is commonly referred to as the *measurement problem*. An unresolved question is – what constitutes an observation? Does a wavefunction collapse only when the observation is carried out by a conscious being, e.g. a human? This was the view of

Eugene Wigner and John von Neumann, two highly eminent quantum physicists. Since we find it to be that case that where there is more than one observer they agree on the result, presumably the first person to carry out the observation brings about reduction of the state vector, so that their colleagues see what they have seen.

Alternatively, does the wavefunction also collapse when the quantum system is 'observed', whatever this might mean, by inanimate detection and recording apparatus? In view of its failure to resolve this dilemma it seems that quantum mechanics is incomplete. A number of attempts have been made to fix the problem by tinkering with the Schrödinger equation, e.g. by adding extra, stochastic, terms which would, in a probabilistic manner, make the wavefunction reduce to a form corresponding to a definite value of one particular measurable quantity.[19] A promising escape route from this impasse has been opened up by *quantum decoherence* theory.[13,20] In essence, this says that while a superposition of the wavefunction of all possible outcomes of an event is initially established, it very quickly disappears, or collapses, as a consequence of interaction with the surrounding environment. The original superposed state is said to be *coherent*, so the process which brings about a collapse of the quantum superposition to a single definite state is referred to as *decoherence*. Calculations indicate that tiny environmental interactions, such as collisions with air molecules or radiation photons, are enough to trigger the collapse. Thus, according to quantum decoherence theory, when an event with more than one possible outcome occurs, reduction of the state vector to a state corresponding to just one outcome rapidly takes place spontaneously whether or not the event is observed. It is not, however, universally accepted that decoherence theory in fact solves the measurement problem.[16]

The quantum mechanical dilemma concerning the apparently necessary role of an observer is reminiscent of that strand of idealist philosophy which holds that nothing exists unless it is actually observed. This view was summed up by an Oxford undergraduate in the 1930s, with a limerick about a tree in an Oxford college quadrangle:

There was a young man who said 'God
Must think it exceedingly odd
That the juniper tree
Just ceases to be
When there's no one about in the quad'.

Another undergraduate, Ronald Knox, who later became well-known as a writer and a Catholic cleric, responded with:

Dear Sir, it is not at all odd
I am always about in the quad.
Thus the juniper tree
Never ceases to be
Since observed by, yours faithfully, God.

In his books *The Mystery of the Quantum World*[18] and *Conscious Mind in the Physical World*,[19] theoretical physicist Euan Squires reminds us that there is an all-embracing solution of the measurement problem, which retains the crucial role of the conscious observer in bringing about collapse of the wavefunction. This is the hypothesis that there is an omnipresent conscious entity which observes every event in the universe and thus ensures that the wavefunctions collapse. The 'God as the conscious observer' theory has the merit that it would explain how the physical universe was able to proceed through all its necessary stages of development even before human (or non-human) conscious observers had appeared on the scene. While this proposal is not likely to be taken up with enthusiasm by the quantum physics community it can, if all else fails, be retained as the 'hypothesis of last resort'.

Quantum Mechanics and Reality

In summary, then, just what does science, through the medium of its most successful modern theory – quantum mechanics – tell us about the

fundamental nature of physical reality? It tells us that at the level of quantum objects, there is no strict cause and effect such as we observe to exist in the macroscopic world. Thus while we may say that any particular type of event, such as radioactive decay of certain atoms, has a certain reproducible (in a statistical sense) propensity to occur, we cannot say that any individual event had a specific cause which made it happen just when it did.

Not only, according to quantum mechanics, can we not speak of a quantum object being subject to strict causality, but we cannot even suppose it to have a precise location in space. All we can do is to use the wavefunction to calculate, for some region of space where the object might be thought to be located, the variation with position of the probability of it being found if a position measurement is actually made. The implication is not merely that the object is hard to find, but that it really exists in some curious spread-out non-localized manner. A related principle is that every quantum object is both a form of wave energy, and a particle. Even seemingly unambiguously particulate entities such as electrons, or large molecules, when they encounter a double slit in a screen behave as if they pass through both slits at once and show wave-like interference on the other side.

When a quantum object is presented with a choice of possibilities, such as a photon encountering a half-silvered mirror at which reflection or transmission may occur, quantum mechanics says that both possibilities happen and the state of the quantum object is described by a superposition of the two possible states: the wavefunction after the encounter is the sum of the two different wavefunctions. When we make a measurement on the quantum object it is, as it were, forced to choose one of the two options, and our measurements – with apparatus at the macroscopic scale – will find only one of the possibilities to have occurred.

When there are two particles which, as a result of some interaction, or perhaps common origin in the decay of an unstable atom, have a certain necessary relationship in their values of some property such as spin or polarization, quantum mechanics says that they are in an entangled state. This means that when a measurement is made of the property in one of

the particles, forcing it to exhibit one specific value out of the possibilities open to it, this property in the other particle instantaneously acquires the complementary value even though no measurement has been made on it and no matter how far apart the particles may be.

In all situations in which a measurement is made on a quantum system, out of the range of possibilities which simultaneously pre-exist in proportions prescribed by the wavefunction, only one is actually found. The wave function collapses.

It may well be that the intelligent reader, however well-disposed towards science, finds the picture of ultimate physical reality presented by quantum mechanics too bizarre, too counterintuitive, too contrary to everyday experience to accept. If this is your reaction you need not feel guilty about it – you are in good company. Some of the founding fathers of quantum mechanics – Einstein, Schrödinger, de Broglie – were deeply unhappy about some of these implications that began to emerge as the subject they had helped to create took on a life of its own in the 1920s and 1930s.

Hidden Variables – The Pilot Wave

Is there an alternative? If you really cannot accept the quantum mechanical picture of reality, are there other views of the ultimate nature of things which are at least consistent with science, that you can opt for instead? Indeed there are, although it must be said that none of them command a substantial following among scientists. I refer here to *hidden-variable theories* of quantum phenomena. Earlier in this chapter we discussed the possibility that the apparent indeterminism of events at the quantum level might in fact be due to entirely deterministic processes acting upon hidden variables within the system. The best-known hidden-variable theory is that first proposed in 1926 by French physicist Louis de Broglie (1892–1987). His earlier, very important, contribution to quantum mechanics was to propose that just as electromagnetic wave energy has particulate properties,

then particles such as electrons should also have associated wave-like properties, and to work out the implications of this. He became unhappy with the purely probabilistic interpretation of quantum phenomena taking hold at that time, and suggested that any given quantum object, such as a particle, consists of two entities – an actual particle which at any time has a real location, and an associated wave field (sometimes called the *pilot wave*) which in some way guides the particle. It is the pilot wave, not the particle itself, which undergoes interference effects in double-slit experiments. Thus the wave-particle duality arises because both waves and particles really exist, but are linked in some intimate manner.

A fully worked out version of the pilot-wave theory was developed by David Bohm in the 1950s, and is usually referred to as the Bohm–de Broglie theory. According to this, what the wavefunction of a system, calculated with the Schrödinger equation, describes is a *quantum force* or *quantum potential*, which is one more force acting on the particle and which, in combination with any other influences, such as electrical or magnetic fields, determines the behaviour of the particle. In terms of conceptual palatability, the Bohm–de Broglie theory has the advantage that it says particles are real and at any given time have a definite location, and their behaviour is entirely deterministic. It does not, however, remove the necessity for instantaneous interaction between pairs of entangled particles separated by great distances, when a measurement is carried out on one member of the pair. So restoration of common-sense reality is not quite complete.

Why are hidden-variable theories such as this not highly regarded by most quantum physicists? The main reason, as put by theoretical physicist Euan Squires in *The Mystery of the Quantum World* is that 'although hidden variable theories are *possible*, they are, in comparison to quantum theory, extremely complicated and messy. There are other more technical reasons. The hypothesized quantum force appears to be very different in its properties from the other forces that exist in nature, and in particular has the remarkable property of being unaffected by distance.

Nevertheless, although unpopular with the theoretical physics community, it must be said that hidden-variable theories of the Bohm–de Broglie type are still scientifically permissible, so those who resolutely hold to the view that reality must be concrete and deterministic, and in principle if not in practice precisely describable, may seek refuge there if they like.

From within the ramparts of our chosen hidden-variable theory we may congratulate ourselves that we have got determinism back again. Vive Laplace! But … not so fast. If we have got it back at all (and this looks very doubtful) it is only at the quantum level. Up at the macroscopic level, where we actually live and where it was once assumed that determinism obviously prevailed, things are now looking quite uncertain.

Indeterminism in the Macroscopic World

Over the last few decades it has slowly dawned upon the world of physics that the macroscopic world is not deterministic after all. Laplace was wrong. How this extraordinary change in the way we see causal relationships in the macroscopic world has come about is a very interesting story in its own right, but we will content ourselves with the principle involved. Consider any physical system which is simultaneously subject to a number of different influences. An example might be the tossed coin discussed earlier. As it rises spinning through the air from the experimenter's thumb, it is steadily losing upward momentum partly by conversion of its kinetic energy to gravitational potential energy, and partly by the air's resistance to movement through it. The coin is slowly losing angular momentum, i.e. spinning more slowly, due to frictional resistance to the rapid turning motion. The precise extent to which these processes occur depends on air viscosity, which is a function of air temperature and which will vary somewhat along the coin's trajectory. On top of all this, the coin's motion will be influenced by any draughts or localized turbulence within the air mass.

You may say, as Laplace would have said, that while all these complications make the precise outcome of any given coin-toss immensely difficult

to predict, nevertheless all the influences are entirely deterministic: the outcome of the toss is fully predetermined, and is in principle calculable by an appropriately programmed and sufficiently powerful computer. In fact, we now know that for many situations this is not the case. These are systems which show the phenomenon of *extreme sensitivity to initial conditions*: this is the situation in which the outcome of a dynamical process can show very substantial changes as a consequence of infinitesimal changes in the values of the physical conditions which influence the process.

That this held for many processes in the physical world was proposed by the great French mathematician Henri Poincaré (1854–1912) at the turn of the 19th century, but the message for some reason did not sink in. It was only rediscovered with the advent of modern electronic computers with which it is possible to carry out calculations on physical processes (*computer modelling*) with the initial conditions specified to an arbitrarily high degree of precision. It was found that when certain physical systems were modelled in the computer their predicted future behaviour depended very sensitively on the precise values given to the input parameters. Two different calculations, differing only minutely in the values of the input parameters, would predict closely similar behaviour of the system over a certain period of time, but eventually divergences appeared which could widen dramatically. Making the values of the input parameters to the two calculations closer and closer, i.e. moving the difference downward in scale by several places of decimals, would merely delay the time at which divergent behaviour appeared. This phenomenon was first observed by US mathematical meteorologist Edward Lorenz, while working on the computer modelling of weather forecasts in 1960, but is now known to apply to many different physical systems.

A particularly striking illustration of how the behaviour of a system can be rendered unpredictable by even the tiniest imaginable uncertainty in initial conditions has been calculated by British theoretical physicist Michael Berry.[4] He invites us to consider a box containing gas molecules moving this way and that, and colliding with each other in the usual

manner. If we know the exact position and velocity of every molecule at a certain time then in principle, according to the Newton/Laplace view of the world, we could calculate the future trajectories of the molecules indefinitely into the future. In order to take gravitational effects into account, we would also need to know the position of every other particle in the universe. Say we in fact know the position of every other particle in the universe, except one – a single electron located somewhere at the observable limit of the universe, about 10 billion light-years away, but we do not know exactly where. Berry then asks – after how many molecular collisions in the box of gas will the unthinkably small uncertainty in the motion of the gas molecules caused by the uncertainty in the location of the electron, show up as serious errors in prediction of the gas molecule behaviour, e.g. an error of 90° in the predicted angle of a molecule after a collision? The answer is that such errors would show up after only about 50 collisions, corresponding to an elapsed time of only a tiny fraction of a microsecond. A corollary of this sort of analysis is that the uncertainty which is an inherent characteristic of quantum systems could in certain circumstances propagate up to the macroscopic level to influence the outcome of events in a system which exhibits extreme sensitivity to initial conditions.

Physical systems which show this kind of behaviour are referred to as *chaotic*, and the general name for the phenomenon is *chaos*. Fortunately for our understanding and management of the world, not all systems are chaotic. The planets will continue to circle around the Sun, and the Moon around the Earth, in an entirely predictable manner for a very long time. While the details of the weather beyond the next few days may be inherently impossible to predict, nevertheless we may be sure that every year the relative cool of winter will give way to the warmth of spring and summer. We may be quite unable to predict whether the next toss of the coin will come up heads or tails, but we can with some confidence at least predict within narrow limits where on the floor the coin will fall. The fact remains, however, that some chaotic systems up here in the macroscopic world are

so unpredictable that, as physicist Paul Davies puts it,[7] their behaviour could not be predicted even if the entire universe was used to try and compute it. Thus macroscopic indeterminism, although of a less spooky and philosophically problematic nature than quantum indeterminism, is nevertheless absolutely real.

Quantum Mechanics and Gravity

Quantum mechanics describes very accurately and completely the behaviour of ultramicroscopic objects – molecules, atoms, subatomic particles, photons. Newton's theory of gravitation and Einstein's general theory of relativity very satisfactorily describe the behaviour of macroscopic entities – galaxies, gas clouds, stars, planets, freely falling bodies. There is, however, no theory within physics which in a natural way combines quantum mechanics and gravity. We can simply accept that each of these theories is the appropriate one for the level of physical reality which it addresses, and stop worrying about the fact that there is no obvious connection between them. This is the approach recommended by Freeman Dyson.[10] Some physicists find this unacceptable, and for them the Holy Grail is an all-encompassing theory of *quantum gravity* which would combine quantum mechanics with general relativity. One attempt to achieve this makes use of what was originally known as *string theory*, but is now sometimes referred to as *M-theory*. This began around 1970 and has become a very active field, soaking up the energy of a significant proportion of the theoretical physics community. It hypothesizes that the various particle types that have been observed are not points but very small ($\sim 10^{-35}$ m) strings which vibrate at specific frequencies. The strings can be closed loops or open. Different vibrational modes of the string correspond to different particle types, e.g. in one mode the string may be an electron, in another it may be a photon. There is even one postulated mode corresponding to the *graviton*, a hypothetical particle which carries the force of gravity, raising the exciting possibility of combining gravity and quantum mechanics in one theory.

The idea of all the different particle types being just different modes of vibration of a fundamental string-like entity is attractively simple. Unfortunately, the simplicity ends there. Mathematically, it is very complex. Furthermore, to make string theory work it has been found necessary to hypothesize that space-time consists not just of four (three dimensions of space, one of time), but 10, 11 or possibly 26, dimensions, depending upon which version of the theory is favoured. It is supposed that we cannot perceive the extra dimensions because they are very tightly curled up. The analogy often used is that of a garden hose lying on a flat surface. From a distance it looks like a line on a two-dimensional flat plane. Close up, however, we find a third dimension curled up inside it. A worrying feature of the theory is that it throws up too many permissible possibilities. For example, it can apparently generate 10^{500} models of the universe.[5] An even more serious problem is that although string theory has existed for some decades it has so far failed to give rise to predictions that can be experimentally tested. A theory that cannot be tested by experiment is not part of science, although it could still be true. We can only hope that the string theorists will eventually come up with a prediction that can actually be put to the test.

The fundamental physical constants

The equations of physics are statements which tell us, in compact mathematical form, precisely what way certain physical properties of the real world are related to or determined by, and can be calculated from, other physical properties. They can, for example, describe static relationships. a well-known example is Einstein's equation:

$$E = mc^2$$

which enables us to calculate the rest-mass energy of a body with mass m, c being the velocity of light. They can also describe how a feature of a physical

system can vary with time. For example the velocity v of a body which has been falling freely for time t seconds, just above the Earth's surface is:

$$v = \frac{M_E G}{R_E^2} t$$

where M_E and R_E are the mass and radius of the Earth respectively, and G is the gravitation constant.

Other equations can specify how a particular physical property can vary with position. For example, equation (6) in the *Schrödinger wave equation* box:

$$P(x, x + dx) = \frac{2}{L} \sin^2 \left[\left(\frac{\sqrt{2mE}}{\hbar} \right) x \right] dx$$

expresses the probability of a particle which is confined within length L, being found at distance x. This probability, $P(x, x + dx)$, depends on L, the mass and energy of the particle (m and E), and on $\hbar$ (Planck's constant/2π), together with x.

We observe that each of these equations contains, as well as variables specific to the system in question (m in the first equation, M_E and R_E in the second, L, m and E in the third), a physical constant (c in the first, G in the second, h in the third) which is quite independent of the system. Other such constants include e, the elementary charge on the electron; α, the *fine-structure* constant; ε_0, the *electric* constant; and m_p/m_e, the ratio of the mass of a proton to an electron. These and a variety of other physical constants are fundamental properties of our universe. We do not know why they possess the particular values that they have. We do know that for there to be a physical universe with galaxies, stars, planets, the full range of chemical elements and finally living creatures, there is very little leeway for variation in the numerical values of these constants (see the discussion of the *anthropic principle* in Chapter 8). They come as a package deal; without that whole package, precisely as supplied, we would not be here.

THE SCHRÖDINGER WAVE EQUATION

For a particle of mass *m*, moving in a region where its potential energy as a function of position (*r*) and time (*t*) is $V(r,t)$, Erwin Schrödinger in 1926 postulated that the following relationship holds:

$$i\hbar\frac{\partial\psi(r,t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi(r,t) + V(r,t)\psi(r,t) \qquad (1)$$

where $\hbar$ is Planck's constant (*h*) divided by 2π, ∇^2 is the Laplacean operator $(\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2)$, i is $\sqrt{(-1)}$ and ψ is the wave-function. It can be shown that if the potential field, and therefore the potential energy, is static (does not vary with time), then equation (1) can be replaced by the relationship:

$$E\psi(r) = -\frac{\hbar^2}{2m}\nabla^2\psi(r) + V(r)\psi(r) \qquad (2)$$

where *E* is the total energy of the particle. This is known as the *time-independent* form of the Schrödinger equation. If, for simplicity, we confine our attention to a particle moving in one dimension (*x*), then equation (2) can be rewritten:

$$E\psi(x) = -\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) \qquad (3)$$

For a particle moving freely in the absence of any potential field, $V(x)$ is zero and so we obtain:

$$E\psi(x) = -\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} \qquad (4)$$

From this we can proceed at long last to a simple expression for the wavefunction itself in terms of definable physical quantities because it can be shown that if:

$$\psi(x) = A\sin\left(\frac{\sqrt{2mE}}{\hbar}\right)x \qquad (5)$$

where A is a constant, then $\psi(x)$ satisfies equation (4), i.e. equation (5) represents a solution of equation (4). For a particle confined to a (one-dimensional) box of length L, the constant, A, has the value $\sqrt{(2/L)}$. The probability of the particle being found in the small interval of distance between x and $x + dx$ is equal to the square of the wavefunction times dx:

$$P(x, x + dx) = |\psi(x)|^2\, dx = \frac{2}{L}\sin^2\left[\left(\frac{\sqrt{2mE}}{\hbar}\right)x\right]dx \qquad (6)$$

CHAPTER 4

THE PHYSICAL BASIS OF CONSCIOUS REALITY

My brain, I'll prove the female to my soul,
My soul, the father: and these two beget
A generation of still-breeding thoughts;
And these same thoughts, people this world
In humours, like the people of this world,
For no thought is contented.

WILLIAM SHAKESPEARE, *Richard II*

Nobody has the slightest idea how anything material could be conscious. Nobody even knows what it would be like to have the slightest idea about how anything material could be conscious. So much for the philosophy of consciousness.

JERRY FODOR (PROFESSOR OF PHILOSOPHY, COGNITIVE SCIENCE)

THERE ARE TWO KINDS of reality. There is the reality of the hypothesized physical world, which we explored in the last chapter. But before this comes the reality of our conscious minds, the only reality to which we have direct access. It is within the conscious mind that human sensibility has its being. Let us now turn the searchlight of science on the very seat of sensibility and ask if it can tell us anything about the nature of consciousness itself. In particular, we shall consider to what extent the quantum mechanical view of ultimate physical reality, as outlined earlier, may be relevant to

the processes underlying or accompanying conscious thought. The question we address here is – how does conscious thought arise in the physical world?

Idealist and realist philosophies

What do we mean by 'consciousness'? We mean the inner world of sensation and emotion that each of us possesses – colour, taste, smell, sound, touch, joy, fear, anxiety, love, beauty – all the things that, in common parlance, 'go on in our minds'. I know that I have this private inner world – it is in fact the only world of which I have direct certain knowledge. Does anything else exist? I choose to believe so. Like most people, I am prepared to make the great act of faith of assuming that corresponding to my sense impressions (the conscious sensations mediated by the five senses), there exists, independently of my conscious mind, a world which we call the physical world.

We may say that surely it is obvious that the physical world exists: can we not see it, feel it, smell it, taste it? Alas, no. That a physical world exists independently of our internal sensations is simply an unsupported assertion. All we have is the content of consciousness. The red colour of that rose in the garden exists only in your mind. If you look to science for help it will simply talk in terms of the predominant wavelengths in the reflected light. You pick the rose and hold it to your nose. That haunting fragrance? Ask science again, and this time you will be told of volatile terpenoids binding to specific protein receptors in the nose and of electrical signals along the olfactory nerve activating certain neurons in part of your brain. But whence *that* particular sensation? Or indeed any conscious sensation?

But surely no-one actually believes that the physical world does not exist, do they? Yes, they do. In the history of human thought there has been a whole family of philosophies, commonly referred to as *idealist*, which hold that the external material world has no separate existence and is a creation

of the mind. The best-known proponent of this view in the Western world was the 18th century philosopher George Berkeley (1685–1753). A less extreme form of idealism was propounded by Immanuel Kant (1724–1804) who, while apparently acknowledging the possible existence of an external world, believed that the regularities (mathematical and scientific laws etc.) which we believe we recognize in the world are not objective features of things in themselves but are structures imposed by the mind. Later idealist philosophers included Hegel (1770–1831) and Schopenhauer (1788–1860).

I mention idealist philosophy not to persuade you to adopt it but to make the point that unfortunately it is not a simple and obvious truth that the physical world exists and that our conscious sensations give us direct access to it. Rational people have sometimes argued otherwise, and it must be realized that the idealist world view, while it cannot be proven, can be presented in a coherent and logically unassailable form. It is said that Dr Johnson's response to Berkeley was to kick a stone in the street and say 'I refute him thus!' Regrettably, such refutation is not actually valid.

The opposite philosophical view to idealism is that of realism, which may be defined as the belief that the physical world exists independently of human observation of it. This is the position of the great majority of practising scientists. It is certainly my own. Indeed the whole scientific research enterprise would lose its point if we believed that there was not really anything out there to be studied. It must, however, be acknowledged that quantum mechanics (Chapter 3) has muddied the waters as far as a simple realist interpretation of the world is concerned, but not to the extent that realism needs to be abandoned altogether.

In this discussion, therefore, we shall assume that the realist position is correct. We shall also assume that the accumulated corpus of scientific knowledge at the beginning of the third millennium is in very large part correct; that, while recognizing that there are always areas in dispute at the frontier of all the sciences, nevertheless science presents us with an essentially truthful view of the physical world.

This physical world seems to comprise an enormous range of entities, over all scales of size and complexity, from quarks, electrons, protons, photons, nuclei, stars, galaxies, planets, rocks and rivers through molecules, microbes, plants, animals and human beings, of which I find myself to be one. I find that my conscious mind is permanently (while I live) associated with one particular human body. This human body has receptors for light, sound-waves, heat, mechanical impact, specific chemicals, direction of gravity etc. The responses of these receptors to that part of the hypothesized physical world where 'my' body happens to be located gives rise to that set of conscious experiences that I refer to as my 'sense impressions'.

Among the multifarious and complex sense impressions that we experience there are those that we have learned to identify as 'other people', like ourselves. We each know that we are conscious beings. Can any one of us know for certain that any other human being also has a conscious mind? No, we cannot. We never have access to anyone else's consciousness, nor they to ours. We observe, however, that other human beings are very similar to ourselves, not merely in their physical makeup but also in the way they talk and give other manifestations of what in ourselves would correspond to conscious mental processes. In particular, at least in members of the human species who take an interest in such matters, they seem to recognize the concept of 'consciousness' if we discuss it with them, and will admit to having conscious thought processes of their own. In short, I suggest that just as we should make the leap of faith of assuming the physical world exists, we should also make the even more plausible assumption that the other human beings we find present in that physical world are indeed conscious.

What about animals – are they also conscious? This is a very thorny issue which cannot yet be resolved. We return to it later in this chapter.

Conscious and Non-Conscious Information Processing

Having arrived at the position that consciousness is in some way present in, or at least associated with, the inhabitants of the physical universe that we refer to as human beings, let us now ask the question – what does it do for them? If we regard consciousness as part of our information-processing capability, then how is that capability enhanced by being at least partly carried out at the level of consciousness? After all, we can readily imagine that there might be creatures who look much like us and exhibit complex behaviour patterns, but whose information-processing, like that in a computer, is of an entirely non-conscious nature. Could such creatures be just like us? Or are there certain kinds of information-processing which can only be carried out in a conscious mind?

Until comparatively recently the tacit, and largely unexamined, assumption has been that all human thought processes could in principle be carried out in a non-conscious manner – perhaps by machines. This view has not disappeared, but is no longer confidently held. This can partly be attributed to the failure of the artificial intelligence community to deliver the goods so optimistically promised in the early years. But more fundamentally, there are now serious doubts about whether it is even in principle possible for a machine to replicate human thought processes. Before considering why this is so, let us look for a moment at what goes on in our minds.

To assess the role of consciousness it is important to remember that human beings habitually carry out all kinds of complex and powerful information-processing at a non-conscious level. Consider the following scenario. A choir is singing, and each member has a book with the words and music in front of them. The singers are simultaneously decoding two sets of printed instructions – the text and the notes – and implementing these instructions, not merely as a sequence of articulated words, but words at a precise sequence of different musical pitches. As an example of information-processing, this is quite an impressive achievement.

Are they using their conscious minds to achieve this? The first few times they sing this piece – certainly. The decoding is difficult. They must get it right. They must concentrate. But after a while they get the hang of it. They have not attempted to commit the whole thing to memory and so still need the books. What is going on in their conscious minds now? An anxious concentration on each note? A meticulous attention to the text? Not at all. Their conscious minds are now wandering all over the place. Memories of the events of the day, what they had for dinner, who said what to whom, what will tomorrow hold, now flicker through their minds. Periodically their consciousness will zero in on some part of the piece – a tricky passage, perhaps, that they know they must concentrate on if they are to get it exactly right, and then the mind wanders off again. But all this time, whatever they are consciously thinking about, some part of their cognitive machinery is keeping them on track, making sure that the words and the notes emerge in the correct sequence and with the appropriate simultaneity. That is, throughout their performance the singers are carrying out very sophisticated non-conscious information processing.

Consider another scenario. You are having a cheerful conversation with friends. A witticism – relevant to the topic of discussion – pops up in your mind, and you contribute it to the conversation. Where did it come from? Did you consciously construct it? Usually not: it just suddenly occurred to you. The witticism was created by a non-conscious part of your mental machinery. So where does the conscious mind come in? It is at the conscious level that you will decide whether to actually use the witticism. At lightning speed you assess – is it funny, is it relevant, is it in good taste? And you speak, or refrain, accordingly. Sometimes of course, as we all know to our cost, in our haste to contribute a gem to the conversation we make the wrong decision: it is our conscious mind which makes the mistake, and again our conscious mind which immediately regrets it.

One final example. We are struggling with a problem: a well-defined problem for which in principle we believe there could exist a clear solution. This might be a problem in everyday life – perhaps financial: how we

can restructure our lifestyle so that we can live within our means. Or in the workplace – how to achieve a certain task with the limited personnel available. It may be a mathematical problem – how to solve this equation, or prove that a certain relationship is true. It may be scientific – how to measure this quantity, how to explain that phenomenon in terms of existing theory. We tackle such problems at the conscious level, turning them over constantly in our minds, examining this or that aspect minutely, considering various approaches which we already know about and which seem relevant. Sometimes this works. Sometimes the answer will emerge from precisely such cogitation. But very often we grind to a halt. We can't seem to come up with the solution. And then, quite unexpectedly, an answer will suddenly appear – sometimes when we are thinking hard about the problem, but very often when our mind is wandering freely while we are in the shower or pottering in the garden. We did not construct the solution in a deliberate conscious process: it was created at a non-conscious level and suddenly manifested itself in our consciousness. It seems that the non-conscious part of our cognitive machinery is not only capable of very complex information processing, but can be creative as well.

What role did the conscious mind play? It identified the problem. It launched the search for a solution. It collected and examined the relevant information. When the solution appeared the conscious mind assessed it, concluded that it was indeed a satisfactory solution to the problem, and then began the process of implementing the solution.

Algorithmic and Non-Algorithmic Information Processing

We can now return to the question of whether it is possible in principle for a machine to replicate human thought processes. It is not in dispute that machines – and we will confine the discussion here to modern electronic computers, not the mechanical adding machines of days gone by – can

carry out many of the 'mental' tasks that until recently only human beings could achieve. These range from simple everyday arithmetic to very long and complex calculations which people could do but which would in practice be intolerably tedious to perform.

When a computer carries out a numerical computation or any other task – such as searching for particular patterns in a DNA base sequence, or checking availability of airline seats to fly from Sydney to London on a given date – it does so by using an algorithm. An algorithm is a set of instructions which can be written as a sequence of specified steps or as a flowchart, for carrying out a particular calculational procedure. Computers act in accordance with the programs entered into them. A computer program is an algorithm for carrying out a specific task.

Human beings can also operate with algorithms. Consider the following very simple example. To find the mean temperature for the month of January:

1 Obtain the daily temperature data for January.
2 Add together the recorded minimum values of temperature for every day in January to give *Sum* (*min*).
3 Divide *Sum* (*min*) by the number of days in the month (31) to give the mean minimum temperature, *Mean* (*min*).
4 Add together the recorded maximum values of temperature for every day in January to give *Sum* (*max*).
5 Divide *Sum* (*max*) by the number of days in the month to give the mean maximum temperature, *Mean* (*max*).
6 Add *Mean* (*min*) to *Mean* (*max*) and divide by 2 to give the mean daily temperature for January.
7 In a table, write out the values of mean minimum, mean maximum and mean daily temperatures for January.

The set of instructions, Steps 1–7 above, is an algorithm set out in a form suitable for use by a person. The same algorithm in a form suitable for a computer could be written as follows, using the computer language BASIC:

```
10  DIM TMIN(31),TMAX(31),TMEAN(31)
20  PRINT 'FOR EACH DAY, STARTING WITH JAN 1, ENTER THE
    DAILY MINIMUM FOLLOWED BY THE DAILY MAXIMUM
    TEMPERATURE VALUES'
30  FOR N=1 TO 31
40  INPUT TMIN(N)
50  INPUT TMAX(N)
60  NEXT N
70  MINSUM=0
80  MAXSUM=0
90  FOR N=1 TO 31
100 MINSUM=MINSUM+TMIN(N)
110 MAXSUM=MAXSUM+TMAX(N)
120 NEXT N
130 MEANMIN=MINSUM/31.0
140 MEANMAX=MAXSUM/31.0
150 MEANDAY=(MEANMIN+MEANMAX)/2.0
160 PRINT 'MEAN MINIMUM TEMP. = ',MEANMIN, 'DEGREES C'
170 PRINT 'MEAN MAXIMUM TEMP. = ',MEANMAX, 'DEGREES C'
180 PRINT 'MEAN DAILY TEMP. = ', MEANDAY, 'DEGREES C'
190 END
```

Clearly, both we and the computer can carry out calculations by following any algorithm supplied. How do we differ? We differ because the computer can only arrive at a result by following an algorithm. In contrast, in many instances we can *see* the truth of a proposition whose validity cannot be established by any purely algorithmic procedure. To give a simple example, any of us can readily 'see' that the natural numbers 0, 1, 2, 3, 4, 5 etc. are infinite in extent, i.e. they can just go on for ever. We cannot program a computer to prove this. A very important theorem concerning a fundamental limitation of all rule-based logical/mathematical systems was proved in 1931 by Austrian mathematical logician Kurt Gödel (1906–1978). It can be stated, following the form used by Euan Squires (see below), as:

> *In any finitely describable formal system that is consistent and that is rich enough to contain the rules of arithmetic there are true statements that are not theorems.*

What this means is that in any theory there are true statements which a computer can never prove but which *we can see are true.*[14] The philosopher, John Lucas concluded from Gödel's theorem 'that mechanism is false, that is, that minds cannot be explained as machines'. A 'machine' (and this is the only kind we presently know about) is something that can only reach theorems by following rules. It cannot 'see' a result. It cannot, as we in our conscious minds can, stand apart from the problem, view it from the outside and see that something is true. Theoretical physicist Euan Squires, in his 1990 book *Conscious Mind in the Physical World*, puts it well:

> *We have a concept of truth which the computer has not got, at least not unless, or until, it is conscious. There is something about truth, or perhaps rather I should say about the recognition of truth, that requires conscious mind for its existence; like free-will, or red, or beauty, it is one of the properties of consciousness.*

That there is this fundamental difference between machine computation and human conscious thinking, and in particular that all this necessarily follows from Gödel's theorem, has been argued in great depth in two books by mathematical physicist Roger Penrose. With respect to the perception of mathematical truth, Penrose has this to say:

> *Mathematical truth is not something that we ascertain merely by use of an algorithm. I believe, also, that our consciousness is a crucial ingredient in our comprehension of mathematical truth. We must 'see' the truth of a mathematical argument to be convinced of its validity. This 'seeing' is the very essence of consciousness. It must be present whenever we directly perceive mathematical truth. When we convince ourselves of the validity of Gödel's theorem we not only 'see' it, but by doing so we reveal the very non-algorithmic nature of the 'seeing' process itself.*

Another way of looking at it is to regard computer 'information-processing' (I prefer not to use the word 'thinking' for this activity) as traversing linear tracks on a flat plane, i.e. it is at best two-dimensional. Conscious human thought, by contrast, is three-dimensional – we can stand back from the plane where the logical operations are being carried out and very often see at a glance just where they are going.

In view of these considerations, especially those arising from Gödel's theorem, it seems that we may reasonably conclude that in the absence of consciousness no electronic computer, regardless of its computing capacity, will ever replicate human thought processes. Which then brings us to the crucial question – if machines cannot achieve consciousness, how is it that we can?

The Physicochemical Basis of Brain Function

We noted earlier that each of us finds that our conscious mind is associated with one particular human body, which we come to think of as 'our' body. But where in the body is consciousness located? We naturally tend to assume that it is located in the head, mainly because this is where the receptors for the most important information-containing inputs to consciousness – light and sound – are found. It is natural to think that our mind is just behind our eyes and between our ears. However, even if evolution had given us eyes on the ends of long stalks and ears on our chests we would still have concluded that the conscious mind is linked to the brain. This conclusion has arisen out of centuries of observation of the effects of various kinds of brain injury on mental function, and in recent times from systematic neurophysiological investigation of brain function in various members of the animal kingdom, including humans.

Neurophysiology has been a very active field of research in recent decades, and a great deal is now known about the internal organization of brain function – which parts of the brain are concerned with visual, auditory

and olfactory inputs; where we carry out verbal as opposed to geometrical thinking; where memories are laid down; how we interpret the visual field and so on. But none of this work, fascinating and important though it is, has led to any insight, any hint of an explanation, of why brain processes should be accompanied by conscious thought. At no point, in order to explain what brain anatomists and physiologists observe, has it been necessary to invoke consciousness. And yet it exists. Indeed, as we saw at the beginning of this chapter, it is the only thing any of us can be absolutely sure exists. Perhaps our best hope of understanding the link between brain processes and consciousness is to proceed to the most fundamental level of brain activity, where neurophysiology shades off into physics and chemistry, and it is down this road we shall now travel.

I shall not start with an account of our current understanding of brain structure and function – such can readily be found elsewhere. Suffice it to say that there is general agreement that cognitive activity, both conscious and non-conscious information-processing, is carried out by, or at the very least is intimately linked to, various structures in the brain – the *cerebrum*, *cerebellum*, *hippocampus*, *hypothalamus* etc., in all of which the functioning unit is the individual nerve cell (*neuron*). There are of the order of 100 billion neurons in the human brain. The main cell body of the neuron, containing the nucleus and the biochemical machinery responsible for the usual cellular metabolic and synthetic activity, is referred to as the *soma*. What is distinctive about the neuron is that it is organized to receive and send signals over long distances. From the surface of the neuron a large number of fibres emerge (Fig. 4.1). One of these fibres, the *axon*, is different from all the rest: it is up to several centimetres long, with a certain amount of branching, and it carries the output signal from the neuron. The other fibres, which are shorter, highly branched and present in large numbers per cell, are the *dendrites*, which receive the input signals.

These fibres connect each neuron to hundreds or thousands of other neurons, creating a network of enormous complexity. Where each fibre joins another cell it is terminated by a small swelling known as a *synaptic*

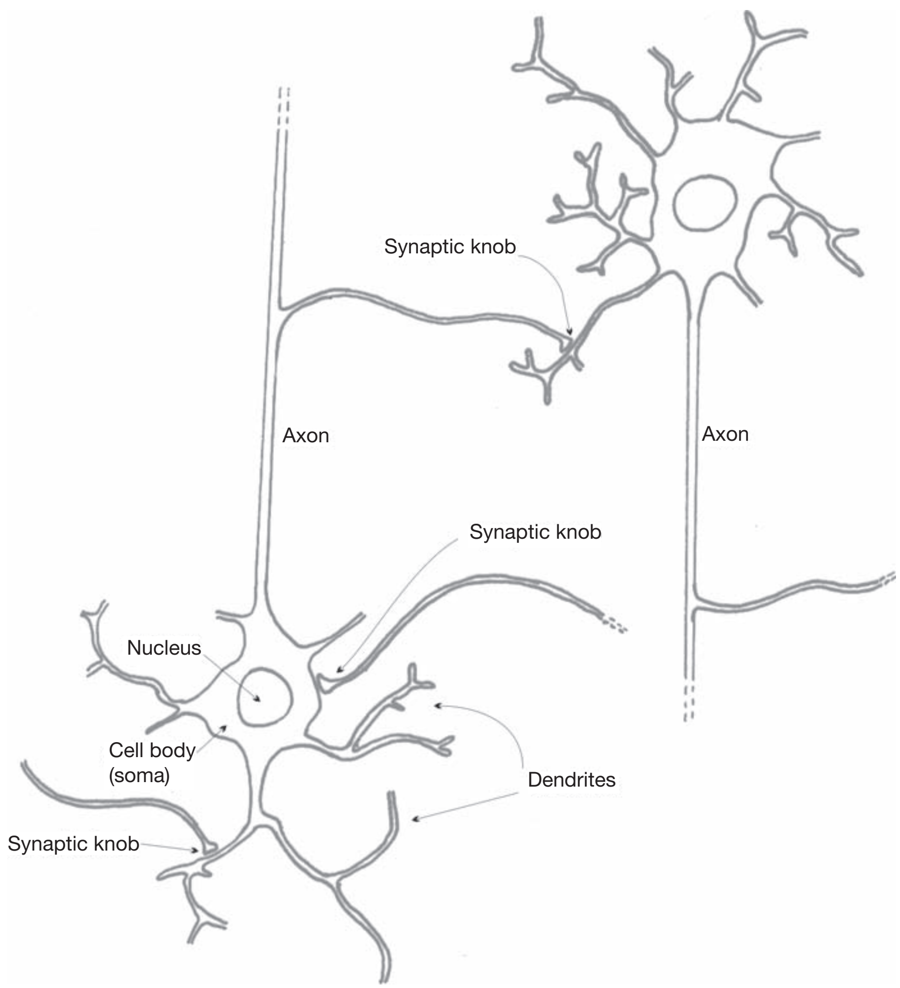


Fig. 4.1 Neurons: general structure and interconnections. From each neuron the axon transmits the outgoing signal to other neurons. These receive the incoming signal through the synaptic knob where it is in direct contact with the surface of the soma or dendrite.

knob (Fig. 4.2); the junction formed by the base of the synaptic knob where it is in contact with the membrane of the dendrite or soma of another cell is known as the synapse. It is across this synapse that signals from one cell to the other must be transmitted.

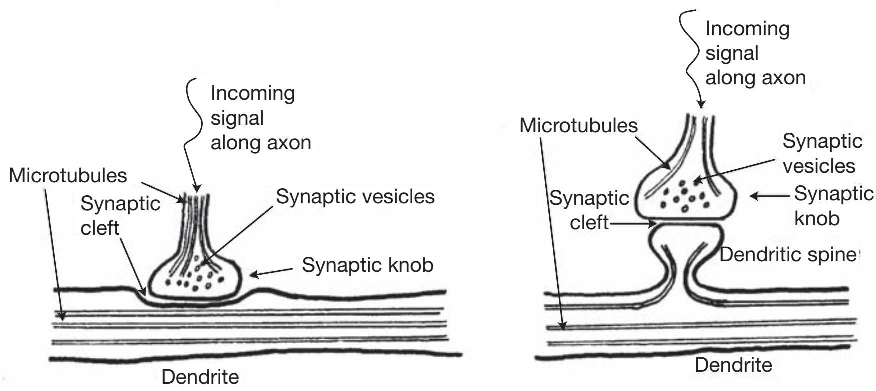


Fig. 4.2 Structure of the synapse, the junction where the signal is transmitted from one neuron to another. The synaptic knob, which is the terminus of the axon fibre originating in one neuron may make contact directly with the side of a dendrite in the receiving neuron, or with a type of protrusion on the dendrite known as a dendritic spine.

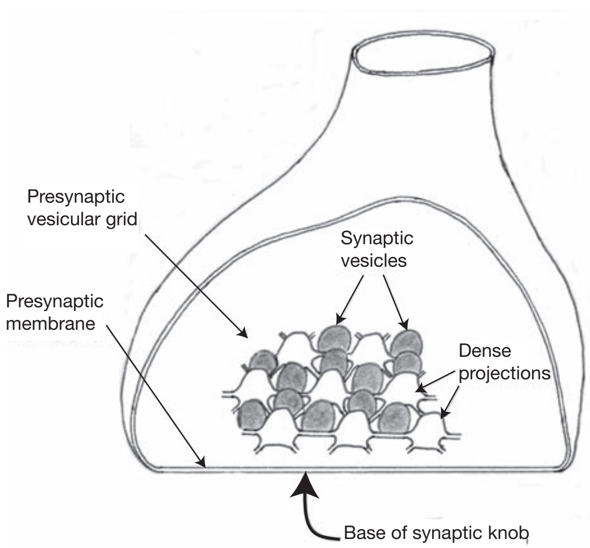


Fig. 4.3 Basal structure of the synaptic knob. On the inner side of the presynaptic membrane at the base of the synaptic knob we see the presynaptic vesicular grid with synaptic vesicles in the regular array of spaces between the dense projections, linked by protein filaments, which together form the structure of the grid. Source: Redrawn from various figures in Eccles, John C. (1994). *How the Self Controls its Brain*. Springer: Berlin.

Transmission of signals along the nerve fibres is electrical. It involves one-way movement of a voltage pulse, the *action potential*, along the cylindrical membrane-bounded fibre, the voltage at any point being determined by the lateral movement of the positively charged sodium and potassium ions between the inside and the outside of the membrane. Transmission of the signal from one cell to the other, across the synapse, however, is chemical and involves movement of molecules, collectively known as *neurotransmitters*. There are believed to be at least 50 different neurotransmitter substances, many of quite simple chemical nature, such as the amino acids, glutamate, aspartate and gamma-amino butyric acid.

The molecules of neurotransmitter substance are contained within tiny membrane-bound packages, the synaptic vesicles, each containing ~5000–10 000 transmitter molecules. These vesicles, hexagonally packed in a two-dimensional grid structure (the *presynaptic vesicular grid*, Fig. 4.3), are attached to the membrane forming the base of the synaptic knob where it abuts the membrane of the other cell. The very narrow space between these two membranes is about 200 Angstrom units (0.02 microns) across, and is referred to as the *synaptic cleft*. When the voltage pulse travelling along the nerve fibre reaches the synaptic knob it depolarizes the membrane, leading to a large influx of calcium ions. The calcium ions combine with the protein *calmodulin*, which then causes a synaptic vesicle to release its content of neurotransmitter molecules into the synaptic cleft. Although there are many (30–50) synaptic vesicles arrayed on the membrane of the synaptic knob, it seems that only one, or none, is discharged by a single nerve impulse. On release, the neurotransmitter molecules rapidly traverse the very short distance across the synaptic cleft and bind to specific protein receptors in the membrane on the other side. The receptor molecules now change their conformation and this causes the opening of channels in the membrane, allowing an influx of certain positively charged ions such as sodium, potassium and calcium. This affects the voltage across the membrane of the receiving neuron, and will consequently affect the likelihood of this neuron transmitting the action potential

arriving from the previous neuron, along its own axon to other neurons. The effect of the neurotransmitters can work two ways, depending on the particular ions they allow into the cell: the probability of an action potential in the receiving neuron may be increased or decreased. Given the possibility of excitation or inhibition at each synapse, given the wide range of neurotransmitter molecules, and given the existence of a number of different neuron types, there is scope for a wide range of functional variability even down at this most fundamental level of operation, in addition to that conferred by the enormous complexity of interconnection of the neuronal network.

The Metaphysics of Consciousness

Now that we have come down to the physicochemical level of brain functioning, can we see any glimmerings of a possible mechanism by which a conscious mind might exist in parallel with, or interact with, the electrical/biochemical processes going on in the active human brain? Before doing so we need to clarify our thoughts about the possibilities with respect to the kind of thing that consciousness might be. This is a matter on which, over the centuries, many different views have been put forward. I will confine myself to summarizing the three main positions. The oldest – originally developed in ancient Greece, but clearly formulated in the modern era first by French mathematician/scientist/philosopher René Descartes (1596–1650) – is that the conscious mind and the physical brain are entities which are quite distinct, but which nevertheless interact. This is referred to as *dualism* or *interactive dualism*. In modern times the dualist view has been powerfully argued by the eminent philosopher of science Sir Karl Popper, and the Nobel Prize-winning neurophysiologist Sir John Eccles. A comprehensive account of the dualist position, with a penetrating critique of opposing views, may be found in their book *The Self and its Brain* (1977), and various books by Eccles, most recently *How the Self*

Controls its Brain (1994). The main technical problem with dualism is to understand how the presumptively non-material conscious mind influences the material brain. We shall return to this later.

At the other extreme we have the purely materialist position which starts from the metaphysical assumption that the physical universe as revealed by the standard methods of physicochemical investigation, is all there is. Nothing else exists. Clearly, to an adherent of this philosophy, consciousness is a serious problem since nowhere in the world of physics and chemistry, neither at the most fundamental level of particle physics and quantum mechanics nor at the higher more complex levels of the neurophysiology/biochemistry of the neuron, or broad-scale information-processing by the whole brain, is consciousness ever observed. The most common response of the ideologically committed materialist to this dilemma is to take refuge in some form of *epiphenomenalism.* This is the view that consciousness is merely something that accompanies, and is entirely determined by, physical brain processes. That these neurophysiological processes (*neuronal correlates of consciousness*) and conscious thought are the 'outside' and the 'inside' of the *same* event. The main problem with epiphenomenalism is that it has no explanatory value. We are still left with the question – but *why* should these particular physical processes be linked with, or give rise to, conscious thoughts? How indeed can we ever persuade ourselves (and why should we even try) that two things – conscious thought and neurophysiological processes – which are palpably quite different in nature, are in fact exactly the *same* thing?

This is, however, not the only problem. It is an essential part of materialist epiphenomenalism that the conscious processes have no causative power; that the thoughts are merely an expression of, and are helplessly carried along by, the biochemical/electrical processes going on in the neural circuits. This being so, then consciousness cannot possibly have any role in evolution, since the causally ineffective conscious thought processes cannot influence the neurophysiological processes going on in the brain as they grind inexorably and deterministically along in accordance with

the laws of physics and chemistry. And yet it is a commonplace in the materialistic world view that consciousness is just one more of those useful faculties that arose in our evolutionary history by natural selection, because of advantages that it conferred.

The third metaphysical position is that of *panpsychism*, which claims that everything in the world, down to rocks or elementary particles such as electrons and protons, is imbued to some extent with consciousness. Something akin to panpsychism seems to be implicit in the 'process philosophy' of the eminent philosopher and mathematician Alfred North Whitehead (1861–1947), and has been taken up in recent times by the Australian biologist/theologian Charles Birch. Proponents of this view would not claim that coherent conscious thought processes are taking place within a rock, but rather that with every object there is an inner 'psychical' aspect as well as the outer material aspect. We would suppose that any psychic aspect, or consciousness, associated with a rock would be of an extremely dim and static nature! As the scale of complexity of the object or system increases, especially once we get into the animal kingdom, so presumably the complexity of the associated consciousness increases, until eventually it reaches its full flowering in the self-conscious thought processes associated with the human brain.

There are many problems with panpsychism as an explanatory hypothesis. One is that nowhere, in all the properties of matter that physics and chemistry have now revealed in such detail, is there anything that seems to correspond to this primeval form of consciousness with which matter is claimed to be endowed. Another is that it provides no mechanism for individual consciousness. If consciousness is merely a property of matter, what is special about the consciousness that seems to be associated with my brain? Does my right leg also have a consciousness? My toenail? And so on. I do not wish to imply that, because of these difficulties, panpsychism can simply be disregarded. What I do say is that to develop it into a theory of mind that deserves serious attention it needs a lot more work.

INTERACTION BETWEEN THE CONSCIOUS MIND AND THE BRAIN

After that brief survey of theories of the nature of consciousness, let us return to the question of the possible mechanism by which a conscious mind might interact with the electrical/biochemical processes going on in the active human brain. The only serious contenders so far invoke quantum mechanics as providing a possible basis for such interaction. This is partly because of a perceived natural affinity between the conscious mind and certain aspects of quantum theory. The physicist Henry Margenau, for example, proposed that the mind may be regarded as a non-material field, analogous to a probability field. Another reason is the realization that quantum indeterminacy might provide a mechanism by which conscious mind could influence physical events in the brain. Certain processes at the level of the synapse involve structures so small that quantum mechanics predicts a significant level of indeterminacy as to the outcome. This provides the opportunity for the conscious mind – whether as the separate independent entity envisaged in dualist philosophy, or as the inner psychical aspect of the whole bodily creature, proposed by panpsychism – to influence the course of events.

Two quite specific mechanisms of this type have so far been put forward. One, proposed by neurophysiologist John Eccles in collaboration with quantum physicist Friedrich Beck, takes as its starting-point the surprising fact, discovered[12] by Stephen Redman and his associates in Canberra, that when a voltage pulse arrives at a synaptic knob, although there are 30–50 synaptic vesicles in the presynaptic vesicular grid only one or none is actually discharged, the actual probability of a discharge being about 0.25. Calculations by Beck and Eccles[1] show that each such discharge, or *exocytosis,* is associated with a significant uncertainty of a quantum mechanical nature. They propose 'that the mental intention (the volition) becomes neurally effective by *momentarily increasing the probability of exocytosis* in selected cortical areas'.

The other specific mechanism has been put forward by theoretical physicist Roger Penrose, whose views on the non-computational nature of conscious thought we discussed earlier. We saw in Chapter 3 that, according to quantum mechanics, at the submicroscopic level of molecules, atoms and subatomic particles, the system can be described as consisting of the superposition, or simultaneous existence, of different possible states. When a measurement is carried out on the system – which, since we are macroscopic beings, necessarily involves an amplification up from the submicroscopic to the macroscopic scale, i.e. the scale of classical, not quantum, physics – it is believed that the system jumps from being in its nebulous mixture of possible states to just one of these states. The measuring apparatus records that particular state. The sudden change in the system is referred to by names such as 'collapse of the wavefunction' and 'state vector reduction'. We also saw that recent quantum decoherence theory suggests that quantum superpositions can collapse spontaneously, in the absence of an observer, as a result of system interaction with its surrounding environment.

The state vector reduction phenomenon remains one of the more puzzling areas of theoretical physics. A much greater puzzle is the nature of the relationship between quantum mechanics and gravitation. Penrose proposes that it is in this state vector reduction process – which he refers to as *objective reduction* and believes to be connected with gravitational effects – that some kind of non-computability in physical laws enters. This in turn might provide the physical basis for the non-computable, i.e. non-algorithmic, character of conscious thought. Psychophysiological experiments have shown that conscious thought processes, e.g. a decision to carry out a simple movement, take place with time-scales in the region of 0.2–0.5 seconds, very long times compared to the virtually instantaneous process of state vector reduction as presently understood. The problem therefore is to find a plausible candidate for (presumptively non-computable) objective reduction processes in the brain.

Penrose, with Stuart Hameroff, believes he has found one in the form of the *microtubules*. These are tiny tubes about 0.025 microns in diameter that

are present in all cells but are especially common in nerve cells, where they lie parallel to the axis. The microtubules are built in the form of a regular cylindrical array of molecules of a protein called *tubulin*. Tubulin exists in two forms, α and β, and can undergo conformational change from one to the other. Penrose and Hameroff propose[9] that within the brain there can be a large-scale quantum-coherent state involving many microtubules in many different neurons. A quantum-coherent state is a large quantum system, its elements all acting in concert in such a way that the quantum state vector description applies on a comparatively large scale, not just to an individual atom or particle. A superconductor (a piece of metal whose electrical resistance falls to zero at very low temperature) is an example of a macroscopic quantum coherent system. In the hypothesized quantum-coherent state in the brain, it is envisaged that quantum superpositions of different combinations of tubulin conformation could exist. When the state vector reduction takes place, one pattern of tubulin conformations is selected. This selection takes place in a non-computational way, and when this happens a conscious event takes place. The selected pattern of tubulin conformations (in many neurons) would then affect the reactivity of numerous synapses, and brain functioning would be changed accordingly. If 10^4–10^5 neurons, acting coherently, are involved in a single conscious event, according to Penrose's calculations the time-scale will be of the order of half a second.

A potentially very serious criticism of the Penrose–Hameroff theory is that in the warm, wet and noisy milieu of the brain, quantum-coherent states will be much too short-lived to affect brain processes in the manner the theory proposes. Physicist Max Tegmark has published calculations[16] purporting to show that decoherence time for microtubule quantum states must be as low as ~10^{-13} seconds, far too short to affect neuronal processes, which are on the scale of milliseconds. In response, however, the Hameroff group question the validity of Tegmark's theoretical treatment and on the basis of their own calculations[6] arrive at decoherence times in the range relevant to brain processes. This is a field in which the final answers are

certainly not yet in, and we may confidently expect interesting new developments in the near future.

The Penrose–Hameroff theory of course does not provide an explanation in physical terms of the existence of such a thing as conscious experience. What it does is suggest particular structures and particular processes with which non-algorithmic information-processing might possibly be associated. As it stands it is compatible with any of the three major philosophical positions – dualism, panpsychism or epiphenomenalism – although Penrose himself does not favour the dualist view. Before leaving the possible role of quantum coherence as the link between brain processes and consciousness, we may wonder whether the presynaptic vesicular grid, with its regular paracrystalline structure, is another candidate for participation in a quantum-coherent state. This might provide an alternative way in which the release of the single synaptic vesicle on arrival of the voltage pulse at the synaptic knob could be subject to quantum uncertainty, in the mechanism of Eccles and Beck.

Quantum physicist Henry Stapp has proposed[15] a quantum mechanical mechanism for the interaction between the physical brain and the conscious mind which is less specific and biochemically detailed than those of Eccles or Penrose, but which can be regarded as in principle encompassing either of these, or a number of other possibilities. We saw in Chapter 3 that until it is actually observed, a quantum system must be described by a wavefunction which is a superposition of a number of distinct wavefunctions, each corresponding to a different possible state of the system. When the act of observation is carried out, the wavefunction 'collapses' and the system ends up in just one of the possible states. Or, in the quantum mechanical version of reality proposed by Heisenberg and Popper, the physical universe evolves with time in the manner prescribed by the Schrödinger wave equation, but this evolution is punctuated at certain times by actual events, or *quantum jumps*. What the quantum mechanical theory specifies is the propensity for any given outcome when a 'detection event' actually occurs. Putting it yet another way, what is evolv-

ing with time is the probability distribution of the different possible outcomes of a detection event.

Stapp proposes that brain processes also evolve with time in this manner, so that at any given time there will exist a superposition of macroscopically distinctive states such as, say, both the activation and the non-activation of the neural activities corresponding to the intention to raise an arm. The conscious sensation of deciding to raise or not raise the arm corresponds to a 'detection event' in which just one of these macroscopically distinct neural activities is actualized. As Stapp puts it, 'Human conscious events are assumed to be the *feels* [my italics] of these top-level events which actualize *macroscopic* patterns of neural activity'. (By 'feel' in this context, Stapp means conscious experience.) While Stapp's approach does not provide a physical explanation of why there should be such a thing as consciousness, his proposal that each event in the conscious mind is directly linked to a selection, by collapse of the wavefunction, between previously superposed (in a quantum mechanical sense) but distinctive macroscopic patterns of neural activity, is an interesting and potentially fruitful hypothesis on how quantum mechanics may be involved in the mind-brain phenomenon. It is not yet clear, however, the extent to which this kind of explanation is affected by the implication from quantum-decoherence theory (Chapter 3) that quantum-coherent states tend to undergo very rapid spontaneous collapse due to interaction with their environment.

In the light of the various proposals for the mechanism of brain-mind interaction, how should we now assess the different philosophical standpoints on this subject? In my own view, the advantage of a dualist hypothesis of the Descartes/Eccles/Popper type is that in principle it provides a complete explanation. It says that there are two levels, types, forms, whatever, of reality – the physical and the conscious/mental (Popper's Worlds 1 and 2), proposes that they can connect, and offers a suggestion as to how. The problem with the physicalist type of explanation, even those of the most liberal and open type which acknowledge the reality and distinct functional capability of conscious thought – here I refer particularly to the

proposals of Roger Penrose – is that even if for the sake of argument we accept that something along the lines of quantum collapse in microtubules in neurons may in some way provide a mechanism for non-algorithmic thinking, it is still unclear as to how this should give rise to anything so peculiar, and utterly alien to the physical world, as consciousness.

ANIMAL CONSCIOUSNESS

Finally, let us turn to the vexed question of animal consciousness. Is *Homo sapiens* the only conscious species on this planet, or does consciousness exist in other organisms too. If so, how far into the animal and perhaps even the plant kingdoms does it extend? Let us acknowledge at once that we do not know, and it is entirely likely that we shall never know. We are each conscious and we believe that other human beings are conscious, partly because of their similarity to us but more particularly because we can communicate with them and they tell us that they are conscious. We cannot ask animals if they are conscious, so the most we can do is to see if we can legitimately, or at least plausibly, infer the presence of consciousness from their behaviour.

In the case of the higher animals, which sometimes show behaviour that in a human being would be regarded as expressive of an emotion – anger, fear, shame, devotion – our natural reaction is to assume that the animal is consciously experiencing these emotions. Or when we see animals – squirrels, crows, raccoons – showing ingenious and not merely instinctive behaviour in gaining access to food supplies, for example, it is tempting to suppose that some form of conscious, if not verbal, cogitation has enabled them to do this. However, heartwarming though all these suppositions are, if we wish to be completely rigorous (and I am not suggesting that we necessarily should) we should abstain from attributing consciousness in these cases. We should seek to explain the observed behaviours in terms of the complex and sophisticated internal information-processing of which

the animals' brains are capable. External expression of emotions such as anger and fear can arguably be explained as having survival value in animal interactions, and so could have been selected for in evolution without the existence of any simultaneous conscious emotions in the animals. Remember too our earlier discussion of just how much non-conscious information-processing, of a very complex nature, goes on in human beings. Solutions to difficult problems are arrived at somewhere in the brain then emerge into consciousness to be examined, and given the *imprimatur,* by the conscious self. And do we not sometimes feel emotion of some kind welling up from somewhere before it finally bursts out into the conscious mind? In *The Self and its Brain* neurophysiologist John Eccles describes how:

> *You can have a decorticate animal with the whole cerebral hemispheres removed and it would still react to pain and show fear and rage, in fact the whole range of basic adversive reactions.*

Thus an animal can exhibit behaviour which we might regard as indicative of the presence of conscious emotion despite the surgical disconnection of that part of its brain with which consciousness is most likely to be associated. Psychologist Clive Wynne, who studies animal learning, cautions against the attribution of consciousness to animals.[17,18] While acknowledging that old-time behaviourism may have imposed excessive constraints on animal psychology, he nevertheless argues that 'the reintroduction of anthropomorphism risks bringing back the dirty bathwater as we rescue the baby'.

These considerations lead us to the conclusion that we cannot with any certainty infer the existence of consciousness within animals simply on the basis of the perceived similarity between some of their behaviour patterns and behaviours in humans which we know to be conscious. On the other hand, if we find the supposition that animal consciousness exists more plausible than the contrary view, then we should feel free to accept it.

How could the matter be resolved? The only possibility I can presently see, and it is a very faint and unpromising one, would be to look in some of

the higher animals for problem-solving behaviour of a type that requires non-algorithmic information-processing, i.e. 'understanding' a problem in a way that a mere computer cannot. Finding clear examples of such problems, with which to test our more intelligent animal species, will not be easy.

Where does animal consciousness stand in relation to the different philosophical positions outlined earlier? As they regard consciousness as solely a manifestation of certain neurophysiological processes in part of the complex human brain, I would assume that epiphenomenalists attribute consciousness only to those animals, if at all, which have complex neural systems and which exhibit the kinds of behaviour that in humans is associated with conscious experience. The adherents of panpsychism have no problem since they believe that some form of consciousness is associated with all matter. For them there is not merely animal consciousness but also plant consciousness, microbial consciousness and even mineral consciousness. I assume that panpsychism would hold that as we move up through the inanimate and animate world to the higher animals and *Homo sapiens*, this consciousness becomes progressively less diffuse, more organized, more self-aware.

The dualist-interactionist philosophy seems to be consistent with a range of views on the existence of consciousness outside the human species. Since its modern version locates the interaction between the conscious mind and the organism within the synapses of the brain neurons, it would regard the possession of a sophisticated neural system of this general type as a minimum, but not necessarily sufficient, requirement for the possession of consciousness. This would exclude all the microbes, all the plants and much of the animal kingdom. Descartes, who made the first clear statement of the dualist-interactionist position in the post-Renaissance era, believed that animals were not conscious. However, Sir John Eccles, the most prominent apologist for this philosophy in modern science, believed that some animal species are indeed conscious, and furthermore that the advantages conferred by the appearance of consciousness played an

important role in evolution. Nothing in the behaviour of reptiles or amphibia, which appears to be entirely instinctive or learned, requires the attribution of consciousness for its explanation. Mammals, however, have a much more varied repertoire. On the basis of his analysis of the organization of the cerebral cortex in the major animal families Eccles proposed[3] that consciousness first appeared when the mammals evolved from the reptiles and is now present only in the mammals, and just possibly also in the birds. When the mammals appeared they had a new structure superimposed on the existing reptilian brain, namely the *neocortex*, and this is what he suggests makes possible the interaction with the world of the conscious mind. He considers it possible that the 'mind world' only came into existence with the evolution of the cerebral cortex in primitive mammals. The neocortex forms the uppermost layer of the cerebral hemispheres. It is only 2–4 mm thick, and in small mammals consists of a smooth layer. In higher mammals it is to varying extent folded and wrinkled, with a corresponding increase in area and volume. This occurs to the highest degree in human beings, where it accounts for about 76% of the brain volume. Birds present something of a problem since some species show behaviour that can plausibly be taken as evidence of conscious experience, but birds do not have a neocortex. On the other hand, what appears to be the most advanced component of the avian brain is a structure in their forebrain, known as the *Wulst*, concerned with vision, and Eccles thinks it just possible that this might confer gleams of consciousness.

All this, of course, is pure surmise. The fact remains that the only biological species we know to be conscious is *Homo sapiens*, and it may be that subtleties of brain structure beyond the mere possession of a neocortex are required before neural processes are accompanied by conscious thought. A selective advantage of consciousness as it exists in human beings is that it enables us to plan ahead. We can imagine, we can mull over, future activities and future situations in our conscious minds and make judgements about how best to act. It is difficult to see how this could be done just with our non-conscious information-processing abilities, even

though we know these can throw up solutions to problems. If we choose to suppose that we are the only conscious species, then the interesting question arises of when consciousness appeared in our evolution. Was *Homo habilis* conscious? *Homo erectus*? Or was the sudden enormous increase in creativity, in technical and cultural complexity, that manifested with the arrival of *Homo sapiens*, the sign that conscious mind had finally appeared on our planet?

One final thought. If animals, or at least the higher species, are indeed conscious then as nature is 'red in tooth and claw' there is a lot of unavoidable suffering going on out there in the natural world. On the other hand, if animals are not conscious beings, then there is no suffering or unhappiness at all in the non-human world and we can stop worrying about it.

Chapter 5

Environmentalism

I wish, O Son of the living God, O ancient, eternal King,
For a hidden little hut in the wilderness that it may be my dwelling.
An all-grey lithe little lark to be by its side,
A clear pool to wash away sins through the grace of the Holy Spirit.
Quite near, a beautiful wood around it on every side,
To nurse many-voiced birds, hiding it with its shelter.
A southern aspect for warmth, a little brook across its floor,
A choice land with many gracious gifts such as be good for every plant.

From *The Hermit's Song*. 9th century Irish poem, translated by Kuno Meyer

There is a pleasure in the pathless woods,
There is a rapture in the lonely shore,
There is society where none intrudes,
By the deep Sea, and music in its roar
I love not Man the less, but Nature more.

Lord Byron, *Childe Harold*

We have seen in earlier chapters how first the cosmos, then our solar system and our planet, then our biosphere with its myriad life-forms, came into being and how we, the human species, capable of conscious rational thought, evolved within it. With our conscious minds, perhaps acquired only recently, we alone among all living species reflect upon our situation. From such reflection there has come an awareness that the quality of the

various environments in which we live can, by both natural and man-made mechanisms, undergo serious deterioration to the great detriment of human happiness, health and even survival. In short, there has arisen the phenomenon of *environmental concern*. Indeed, now that the threat of nuclear war has receded, this is the issue which most seems to exercise the public mind.

The question I wish to address in this chapter is – between the farthest poles of environmental attitudes in our society today, on the one hand the ruthless exploiters who would chop down the last tree on Earth if it would make them a dollar and on the other hand the ecological extremists who view the works of humans with hostility and who seem to want humanity to revert to the hunter-gatherer stage of development – where does the rational, environmentally-aware, scientifically literate person locate themselves? Can science help at all? Is science the answer, or is science the problem? As we shall see, perhaps it is both. To set the scene we shall first ask – what do we mean by *environmental quality*, and how did our concerns about it originate? Let us begin by defining our terms, then look at the history and development of environmental concern in western society today.

The *Shorter Oxford English Dictionary* defines 'environment' as the conditions or influences under which any person or thing lives or is developed. Environmental concern we may therefore take to be concern about the conditions or influences under which we, the human species, and perhaps other species live now, or under which we may find ourselves living in the future. Another term we need to be clear about is 'ecology'. The dictionary defines this as the branch of biology which deals with the mutual relations between organisms and their environment, Thus, as far as the dictionary is concerned, ecology is a branch of science; for professional scientists it still is. People who actually carry out research on the relations between organisms and their environment describe themselves as 'ecologists', and there are innumerable scientific books and journals with the word 'ecology' in their titles. Unfortunately, the word has in recent years been hijacked by the environmental movement. It has become a catch-all

word for environmental concern and everything remotely connected with it. Some people refer to themselves as 'ecologists' when what they actually mean is 'environmental activists'. One of the best-known journals of the environmental movement is published under the name *The Ecologist*. My strong personal preference would be to reserve the name 'ecologist' for professionals working in that scientific field, and to use a term such as 'environmentalist' for environmental activists.

Let us now look more closely at this concept of environmental concern. It has numerous different facets or aspects, and the aspect which excites the strongest resonances differs from one person to another. To some, habitability of the Earth is the most important thing – the Earth as our life-support system, the need to avoid polluting, damaging or depleting our air, land, water and mineral resources. Wildlife conservation has been another very important strand in environmental thinking – what is often perceived as a moral duty to ensure the continued survival of as many of the animal and plant species currently in the biological inventory of this planet as possible (Fig. 5.1). An overlapping but somewhat different concern is wilderness conservation. It overlaps to the extent that the best way to conserve species is to conserve their natural habitat, and in doing so

Fig. 5.1 Wildlife conservation. The giant panda is the best-known of the very many endangered species on our planet and is often used as a symbol by the wildlife conservation movement.

we may also conserve large areas of natural wild ecosystems (Fig. 5.2). Wilderness conservation differs from wildlife conservation to the extent that the concern is primarily with the preservation of wilderness areas to which people can escape or retreat from the modern technological world for spiritual regeneration. There are also some who argue that wilderness areas have intrinsic rights to exist, regardless of the extent to which humanity makes use of them.

To yet another group of people, the main reason for preserving natural ecosystems may be aesthetic. They find forested hills and pristine lakes beautiful to behold (Fig. 5.3), and wish them to be maintained in an undamaged state even if they have no wish to roam about within or on them. The man-made rural environment, such as the patchwork of small ancient fields bounded by hedgerows in parts of northern Europe, is another entirely legitimate object of environmental concern. We must also not forget that for very many people preservation of the urban environment is nearest to their hearts. I am not referring to urban air pollution or crime, but to the preservation and creation of attractive and liveable urban

Fig. 5.2 Wilderness. Mountains and native forests devoid of habitation form a backdrop to Macquarie Harbour on the sparsely populated western coast of Tasmania, Australia.

landscapes (Fig. 5.4) – Georgian terraces in Bath or Dublin, hill villages in Italy, colonial timber houses in New England, Victorian terraces in Melbourne.

Fig. 5.3 The aesthetic dimension of natural ecosystems. The still waters of Lake Hochstetter (South Island, New Zealand) bordered by native forest, with the New Zealand Alps in the distance.

Fig. 5.4 Urban environmental quality. Victorian terraces, Melbourne.

ORIGINS OF ENVIRONMENTAL CONCERN

Having sorted out our terminology and indicated the particular human concerns to which they mainly refer, let us look back on the intellectual ancestry of the environmental movement. Environmentalism had different intellectual roots in different countries. By 'different' I do not mean only that the founding parents were different but that the emphases, the perceived problems, were different.

Concerns about the state of health of the environment, considered part of humanity's life-support system, and early concerns about extinction of species arose in the 18th and 19th centuries as two of the great imperial powers, Britain and France, commenced their colonization of the tropics. Scientists employed by the colonial administrations observed first-hand the degradation brought about by uncontrolled forest clearing and inappropriate agriculture in what had previously been pristine environments.[6] Not only were the scientists worried about the disappearance of species just as they were beginning to make sense of them with the new science of Linnaean taxonomy, but there also arose fear about the possible climatic effects, specifically decrease in rainfall, if forest clearance continued unabated. This argument, together with the observed associated effects of soil erosion and harbour siltation, had an effect on government and gave rise to some measures of forest conservation in the colonies. There was even concern, as long ago as the 1850s, about climatic change due to the changing proportions of oxygen and carbonic acid (CO_2) in the atmosphere. Richard Grove, who has been mainly responsible for elucidating the history of British and French environmental concern in the 18th and 19th centuries, comments that the current awareness of a global environmental threat 'has … consisted almost entirely of a reiteration of a set of ideas that reached its full maturity over a century ago'.[5] Developments in European environmental concern from the late 19th century onwards are described in Anna Bramwell's book[1] on the history of ecology (the movement, not the science). She attributes particular significance in Europe to the ideas of German zoologist Ernst Haeckel (1834–1919). Indeed,

Bramwell attaches particular significance to Germany as a whole – to German philosophers and the particular nature of German society – as the source of the intellectual ferment from which the European 'ecological' movement arose.

Haeckel himself may have been the first to use the word 'ecology'. It does seem that he was responsible for its introduction to the scientific literature. His definition of the term was the accepted scientific one, namely 'the science of relations between organisms and their environment'. But Haeckel's significance goes beyond his invention of the word, and lies more in the development of his philosophical ideas. He explicitly opposed anthropocentrism (the view that the importance of anything is determined by its significance to humankind), and argued that animal species should be seen as equal to humans; that we should, moreover, worship nature and live by nature's laws. He opposed the concept of a split between mind and body, between spirit and matter. His philosophy, which he called *monism*, was initially materialist but in later years developed to the point at which, although he still held that all was one, that one was all spirit. Indeed, his last book was called *God-Nature*. By this stage Haeckel's monism could reasonably be regarded as a form of pantheism, a philosophical belief which holds that God and the universe are identical. Haeckel was a very effective writer and popularizer of science. His views were widely propagated and were very influential, not only in Germany but in translation in the English-speaking world.

One of Haeckel's most important intellectual descendants was one of his students, embryologist and philosopher Hans Driesch, in whose hands monism turned towards *vitalism*. He believed that within all forms of life there resided something he called the *life-force* which had will and purpose, and which manifested in the evolution of life towards ever higher forms. George Bernard Shaw, the famous writer and dramatist, was a great believer in the life-force; in modern times we have seen a suspiciously similar concept reappearing in the *Star Wars* series of science fiction films ('May the Force be with you'!).

Within science the life-force idea was never well received and had little impact. Outside science, however, especially in Germany, it resonated with numerous people of what we might call a 'naturist' disposition. They were people who attached great significance to the natural world and who felt that humanity had become cut off from it by technology, urbanization, intellectualization. A related theme, common in Germany and elsewhere between the wars, was an idealization of peasant life – back to the land and away from the corruption of the cities. A common feeling was that civilization had taken a wrong turn somewhere along the line. I suspect that such feelings are even more widely held today.

Another important influence was Rudolf Steiner, founder of the philosophical movement known as *anthroposophy*. He was also one of the founders of the 'organic' farming movement, particularly one of its more extreme versions, known as 'biodynamic' agriculture. Steiner also believed that the cosmos is imbued with a life-force, that the Earth is alive and the soil is one of its organs. Steiner's biodynamic agriculture movement is one of the sources of the common belief that not only is organic fertilizer better – because the living soil requires humus derived from living matter – but that artificial fertilizer is positively bad. It is an article of faith with adherents of this movement that chemical fertilizers are bad for the soil, and that foods grown with their help are bad for people.

The environmental movement which arose in England in the first four decades of the 20th century shared some common themes with what was going on in Germany, but was less extreme. It derived a lot of its impetus from the perceived destruction of the English countryside as the suburbs proliferated outwards from the cities – a development made possible by technological advances, specifically commuter railway systems and the motor car. A Council for the Preservation of Rural England (now Campaign to Protect Rural England) was established in 1926 in response to those concerns. But there was also, as in Germany, a pervasive sense that civilization had taken a wrong turn, that we had lost touch with nature, that the remedy was to return to the land. The glorification of rural life

can be traced back to the writings of John Ruskin and William Morris in the 19th century.

The novelist D.H. Lawrence was greatly influenced by the ideas of Haeckel and through his writing influenced others, transmitting a certain Germanic flavour of nature mysticism to the otherwise more prosaic English tradition. Earlier literary influences on environmental sensibilities in England can be found in the Romantic poets of the late 18th and early 19th centuries, especially Wordsworth, much of whose poetry reflects his intense feelings for the beauty of nature. Going back much further in time within the British isles, we can find explicit and eloquent appreciation of nature in ancient Irish poetry written more than a thousand years ago.

By the time significant environmental concern was developing in Europe in the 19th century, the European landscape had already been totally transformed by humans. Truly natural wild ecosystems hardly existed in most European countries. In North America, by contrast, much of the landscape was still in its natural state. Native forest and prairie, undammed and unpolluted streams and lakes, were features which many Americans could still directly experience, and whose intrinsic beauty and recreational value they could well appreciate. At the same time, as felling of the forest and ploughing of the prairie proceeded apace, the limited extent and fragile nature of these ecosystems must have been apparent to many. It is not surprising that the first main theme of environmental concern in North America was the preservation of the natural environment, with its associated wildlife. The US was the first country in the world to establish national parks, beginning with Yosemite in 1864. Americans, and indeed the rest of the world, owe a great debt to Henry David Thoreau, George Perkins Marsh, John Muir and others who did so much by their writings, and in some cases political activism, to bring about the climate of opinion which made this possible. Muir was one of the first to put forward the view – now an article of faith with many environmentalists – that wilderness has value in its own right, quite apart from its potential usefulness to people. He can be regarded as one of the founders of the present *biocentric* philosophy, to which we shall return later.

One other very important figure in the US environmental tradition was the ecologist – and I mean an actual professional scientific ecologist – Aldo Leopold, who was influential both before and after the Second World War. There is a much-quoted statement in his book *A Sand County Almanac*, which has become one of the sacred propositions of the environmental movement,

> *'A thing is right when it tends to preserve the integrity, stability and beauty of the ecosystem. It is wrong when it tends otherwise.'*

Leopold's very influential book was published in 1949, which brings us through to the immediate post-war period. During, and for some years after, the Second World War the environment was for obvious reasons put on the back shelf, but as the developed world got back on its feet from the 1950s onward, environmental concern surfaced again.

While there was some intellectual continuity between the environmental movement after the Second World War and that which existed before the war, new themes also emerged. There was a shift in the dominant preoccupations, noticeably so in the 1960s. Looming shortages of natural resources became a particularly important, although by no means new, theme. Pollution of the environment, especially by the new alien compounds of the chemical industry, became a central concern. Worries about industrial pollution in general were not new. Atmospheric pollution by the smokestacks of heavy industry had been recognized as a serious problem since the 19th century: Charles Dickens description of the mythical 'Coketown' in *Hard Times* is a graphic example of Victorian awareness of the noxious environments created by the Industrial Revolution. What was new was the insidious spread and chronic effects of the new and frequently long-lived synthetic chemicals, especially pesticides. Rachel Carson's famous book *Silent Spring* played a crucial role in arousing public concern.

Overpopulation became a very prominent theme and a host of new prophets arose in the wilderness. There were Paul Ehrlich, Barry Commoner and many others telling us that we have sinned, calling upon

us to repent and mend our ways. The most important message from the new generation of environmental prophets was that we are all responsible for the world's environmental problems: that it is the western life-style, with its very high rate of consumption of energy and resources, that constitutes the main threat to environmental quality on this planet. It is not enough to simply seek out individual scapegoats, such as particular polluting industries. The industries are, after all, making things for us.

CULTURAL ORIGINS OF ENVIRONMENTAL DEGRADATION

This question of whom to blame brings us to the question of whether we can identify the historical and philosophical origins of the attitudes and behaviour patterns in our society which have created our present environmental predicament. In the last 40 years there has been a great deal of breast-beating and scapegoat-seeking by that element in the western liberal establishment which loves to castigate their own society and compare it unfavourably with other societies that they imagine existed before we all came along and spoiled everything. The Judaeo-Christian tradition in particular has been held responsible for all our woes because of its anthropocentric nature, its belief that the world was made by God for humankind and that it is therefore entirely proper for us to use it as we see fit.

By selectively quoting particular texts from the Bible, such as God giving humankind dominion over other living things (in Genesis) or telling Noah that humanity must be fruitful and multiply, a sketchy case can be made for blaming the Judaeo-Christian tradition for what western societies have done to the environment. Remembering the scientific perspective of this book, let us put this theory to the test. If it is correct, then other societies which did not share the Judaeo-Christian tradition should not show the environmental destruction and biological extinctions that occurred in the west. When we look at the historical and archaeological records of other societies, however, we find much the same kind of behaviour towards the environment as in the Judaeo-Christian world.

Clive Ponting has described how a number of ancient pre-Christian civilizations, in Mesopotamia, around the Mediterranean and in Central America first thrived and then collapsed due to environmental mismanagement and overpopulation.[9] Long before Christianity, Plato bemoaned the deforestation and soil erosion in Greece. Ancient Chinese writers refer to deforestation, overgrazing and soil erosion.[10] In *The Future Eaters*,[3] Tim Flannery describes the enormous biological impact, especially the rapid extinction of larger game species, that occurred when the Australoid and Polynesian peoples first occupied the lands of the Pacific. In pre-European pre-agricultural Australia, the hunter-gatherer aborigines regularly burned the forest in certain areas to favour the game they preferred, and in doing so totally modified the ecosystem for their own benefit.

To blame the Judaeo-Christian tradition for our environmental dilemma is simply a case of identifying the wrong scapegoat. What we are up against, I suggest, is the natural tendency for any biological species to fully exploit any environment in which it finds itself. This tendency shows up in every human society, whether the Judaeo-Christian societies of western Europe and North America, the Marxist systems which until recently occupied eastern Europe (with much worse environmental problems than the west), or the more recently industrialized societies with Confucian and Buddhist antecedents. The distinctive role of western society in environmental destruction in a technical sense has been to invent so many of the devices, tools and chemicals – internal combustion engine, chainsaw, bulldozer, chlorinated fluorocarbons etc. – which have contributed to the problem. On the philosophical side, there is no case to answer. I shall argue later, however, that when we go looking for solutions, the Judaeo-Christian belief system may be a good place to start.

Environmental concern in the present day

After that admittedly selective overview of the historical and intellectual origins of the environmental movement, let us move forward and examine

the main themes, the prevailing ideologies, evident in present environmental concern. We will particularly consider with how much of this the rational and scientifically literate person can feel at home. Environmental ideologies can be classified in accordance with what they regard as of central importance. Is it human beings? This is anthropocentric ideology. Is it the whole set of living species, all being regarded as of equal value? This is the biocentric view. Is it the total planetary biosphere, regarded as equivalent to a single living organism? This is the Gaian position, which we will deal with in a later chapter. For a comprehensive account of the bewildering multiplicity of modern environmentalist philosophies and movements, consult *Main Currents in Western Environmental Thought* by Peter Hay.[7]

The anthropocentric ideology is straightforward. It is one of enlightened self-interest. Adherents of this position agree that benefit to humankind is the ultimate criterion of environmental quality. We require clean unpolluted air to breathe and water to drink. We must husband those mineral and land resources that we need for our material wellbeing. We seek to maintain significant areas of natural ecosystem for our aesthetic pleasure and recreational use, for scientific study and for the biological resources – new crops, drug plants – which they may yield when we know them better. We endeavour to create urban environments which will be healthy, economically efficient and conducive to social cohesion. To achieve all these aims will require the acquisition and application of scientific knowledge; this is not usually a controversial issue within this ideology. Anthropocentric environmentalists are in general well aware of, and entirely happy with, the role of science in addressing environmental problems. There is no such thing as an anthropocentric environmental movement. It is better described just as a commonly held point of view, I suspect especially among scientists and engineers. There is, however, a recognizable environmental movement which, since it seems to regard the welfare of humankind as paramount, should be included under the anthropocentric heading. That is the *social ecology* movement. Its best-known representative is lapsed Marxist and

former trade union organizer Murray Bookchin. In simple terms, it can be characterized as a left-wing movement with some anarchist sympathies, which attributes most of our environmental and social ills to attitudes of domination towards both nature and other human beings. Bookchin deserves credit for vigorously opposing the anti-people bias of the extremist wing of the ecology movement.

Let us now look at biocentric environmentalists, who are the extreme wing of the ecology movement. Their distinguishing feature is their assertion that the human species has no more value, no more claim to consideration, no more rights, than any other biological species. 'Biocentric equality' is a term often used to characterize this viewpoint. The biocentric standpoint has quite a long history in the environmental movement, certainly as far back as the 19th century. In Europe it can be traced to Ernst Haeckel; in North America to John Muir. The leading thinker of recent times has been Norwegian philosopher Arne Naess. In 1973 he coined the term 'deep ecology' for his version of the biocentric environmental viewpoint, and there is now a deep ecology movement. The intellectual and historical background of that philosophy is ably set out by two of its adherents, Bill Devall and George Sessions, in their book, *Deep Ecology*. Sessions and Naess have formulated a platform for the movement, consisting of eight principles.[2] The first three of these read as follows.

1 The flourishing of human and non-human life on Earth has inherent value. The value of the non-human life-forms is independent of the usefulness of the non-human world for human purposes.
2 The richness and diversity of life-forms are also values in themselves and contribute to the flourishing of human and non-human life on Earth.
3 Humans have no right to reduce this richness and diversity except to satisfy vital needs.

Since the basic tenets of deep ecology, or biocentric equality, are simply assertions of a philosophical or moral nature, which Naess acknowledges were arrived at by intuition, there is not much that science can say about them. It can, however, point out some of the pitfalls that may be encoun-

tered when biocentric egalitarians seek to apply their principles to real-world situations. To what extent, for example, can we usefully talk about all species being of equal value, given that 'species' is an arbitrary and pragmatic mental construct devised by biologists to help them make taxonomic sense of the overwhelming complexity of the world of living organisms? What is a species today might be promoted to being a new genus or relegated to being a mere subspecies tomorrow, in accordance with the currents of taxonomic opinion. Within the science of ecology, it is certainly not the case that all species are regarded as of equal 'ecological' value since in any given ecosystem, some species are much more important to the functioning of that ecosystem than others. In the vast Southern Ocean, for example, krill (the shrimp *Euphausia superba*) is believed to be the base of the whole animal food chain. It would be hard to take seriously an ideology in which krill was considered of no more value than some other invertebrate species present in the same ecosystem but of very rare occurrence, and of no measurable significance to the functioning of the system.

Given that it is a world view based largely on moral and philosophical intuitions, and given that it is concerned with concepts such as 'values' and 'rights' rather than with propositions about the objective nature of the physical or biological worlds, deep ecology can neither be validated or invalidated by science. Nevertheless, given its views on biological species, especially its reluctance to acknowledge any kind of hierarchy or differences in value between species, it must be said that science and deep ecology are not happy bedfellows.

Attitudes to Science Within the Environmental Movement

Finally, I want to address one worrying tendency which is apparent within some elements of the environmental movement, namely a distinct hostility to science and to the scientific approach in general. People of this mind

hold science responsible for most of the environmental problems which beset humankind and for the sense of alienation, of loss of meaning and purpose, that so many people in our society now feel. Furthermore, because of their mistrust of science, this element in the ecological movement do not even believe that science has a role to play in solving the environmental problems which confront us. These attitudes surface periodically in publications such as *The Ecologist* or articles from the extremist end of the organic farming movement. A sentence from an editorial in *The Ecologist* in 1983 gives the flavour:

> *New Gods have emerged – the Gods of Rationality and of Science – and they have banished the animate from the inanimate, with the result that man no longer considers Nature a goddess to be appeased, but rather a subject to be conquered and exploited for his own selfish aims.*

From an article by Edward Goldsmith[4] in the same magazine in 2002:

> *To begin with, the basic assumption of this world view is that, in creating the world, God (or if we prefer, the evolutionary process) somehow did a bad job. Man's mission on this planet, we therefore presume, is to develop technologies that will enable us to redesign the World in accordance with our vastly superior plan.*
>
> *What lends credibility to this infantile world view is that it is the very basis of modern science. And science is seen as the ultimate arbiter of truth. Indeed, science has become very much part of our secular religion.*
>
> *To be seen as scientific is to be seen as right, and therefore beyond question. On the other hand, to be branded as unscientific is to be denied legitimacy and held up to ridicule.*
>
> *In fact the very declarations of our scientists are now imbued with an aura of sanctity, previously reserved for the holy texts of the established religions.'*

Is there anything in this view that science is responsible for our environmental problems? Yes. Let us admit at once that there is. The chlorofluorocarbons that are currently destroying the ozone layer are unquestionably the product of modern science-based industry. As are the chlorinated insecticides and polychlorinated biphenyls which can wreak great havoc as they accumulate in ecosystems. The nuclear power station at Chernobyl which grossly polluted a large area of Europe when it exploded was undoubtedly the product of modern science. Science has certainly contributed to our environmental problems. But only some of these can fairly be laid at science's door. Human beings were digging fossil fuel out of the ground and burning it, and thus initiating the greenhouse effect, before science had anything to say on the subject. Mineral ores were being mined and heavy metal pollution problems were being created long before science came along to explain the underlying chemistry of what was happening. Large agglomerations of people discharging their wastes have been polluting surface water bodies since Roman times without any help from science. Soils were being eroded and irrigation systems becoming salinized long before there was any such thing as agricultural science.

Environmental problems inevitably arise out of the activities of human society, especially where people occur in large numbers. Science has sometimes exacerbated the situation by 'inventing' new environmental problems such as chlorofluorocarbons, but mainly it acts to ameliorate the situation by enabling us to understand underlying mechanisms, and thus to arrive at remedies. If the human waste from Londoners no longer turns the River Thames into an open sewer as it used to, it is because science has shown how to treat those wastes.

The anti-scientific bias of some elements within the environmental movement is ironic as well as worrying. It is ironic because the environmental movement, certainly the movement which has arisen since the Second World War, is very much a product of science. The founding fathers and mothers, the prophets crying in the wilderness, have mostly been scientists. The first signs of incipient environmental problems – the

rising atmospheric CO_2 levels, the hole in the ozone layer – have come from scientific measurements. That environmentalists should turn against science is at the very least inappropriate, some might say ungrateful. The reason why this anti-scientific mood is a worry is that it may delude some people into thinking that our environmental problems can be solved without the help of science. They may even think that the answer is to turn away from science altogether. This is a recipe for disaster. Environmental problems cannot be solved without science. Whether it is the greenhouse effect, loss of the ozone layer, disposal of toxic industrial wastes, pollution of waterways or whatever, it is only by the application of science that solutions will be found. This is not to deny, of course, that there is an immense social dimension to the solution of these problems. Major changes in our patterns of consumption, our life-style, will certainly be required. But there is no escaping the absolute necessity to continue our search for a fundamental scientific understanding of the problems, and the development of technical solutions based on this understanding.

Chapter 6

Science and Environmental Concern

environment, the complex of physical, chemical, and biotic factors that act upon an organism or an ecological community and ultimately determine its form and survival

Encyclopaedia Britannica, 15th edn, vol. 4 (1993)

Man is more the product of his environment than of his genetic endowment. The health of human beings is determined not by their race but by the conditions under which they live

René Dubos, *Man, Medicine and Environment*

Science, as we have seen, is often used as the scapegoat for the world's environmental predicament. In this chapter we confront this accusation directly. We ask – is science to blame for our environmental problems, or does it provide our only chance of solving them? To illustrate the crucially important role that science and science-based technology must play if we are to effectively address our various environmental crises, let us take a look at some of the most serious environmental problems confronting the world today. We shall consider the actual nature of the problems in scientific terms, and how science can be used to solve them.

Global warming

Let us look first at global warming. Why is it that this is now regarded as a serious threat to humanity? It is not that we have simply noticed the world being a bit warmer in the last decade or so. Given the natural variability in climate, that could have happened anyway without portending anything about the future. The crucial information came from a purely scientific study of the composition of the atmosphere. Very accurate measurements carried out at the high-altitude observatory on Mauna Loa in Hawaii since 1958, combined with data for earlier times derived from atmosphere samples trapped in Antarctic ice, have shown that the concentration of carbon dioxide in the atmosphere has been steadily rising for some time, an increase being detectable every year (Fig. 6.1). That atmospheric CO_2 would increase as the result of fossil fuel burning had long been known, but the extent was impossible to predict because of uncertainty about the uptake of CO_2 by terrestrial ecosystems (forests, farmlands, tundra) and the ocean. The Mauna Loa measurements reveal that already, at the beginning of the 21st century, the concentration of CO_2 in the atmosphere has risen to ~370 parts per million (ppm), compared to its pre-industrial level of ~275 ppm. From spectroscopic data on the radiation absorption properties of gases we know that CO_2 absorbs infrared radiation of wavelengths in the 13–18 micron range. We also know that the Earth, warmed by the Sun to an average temperature of about 15°C, or 288 K, and behaving approximately like a black body (as described in Chapter 3), continually radiates infrared energy in a broad waveband centred on about 10 microns, but with plenty of energy in the 13–18 micron range. CO_2 in the atmosphere absorbs some of this re-radiated solar energy before it can escape to space. Any increase in CO_2 concentration therefore leads to increased trapping of the infrared radiation and a consequent warming of the atmosphere. There has already been an increase in average global surface temperature of ~0.6°C since the late 19th century.

So far the science has been clearcut with very little scope for disagreement. From here on, however, when we wish to proceed from increased absorption of radiation to actual effects on climate – how much warming in degrees C, the effects on rainfall, wind, frequency of storms – there is much less certainty because of the very complex and variable nature of the Earth's climatic system. Nevertheless, our only hope is still science. With sophisticated mathematical models of the behaviour of the atmosphere/ocean/land system and computers of great and ever-increasing power, realistic simulations of world climate at varying CO_2 levels can be made. With continual further refinement these are our best bet for predicting our climatic future.

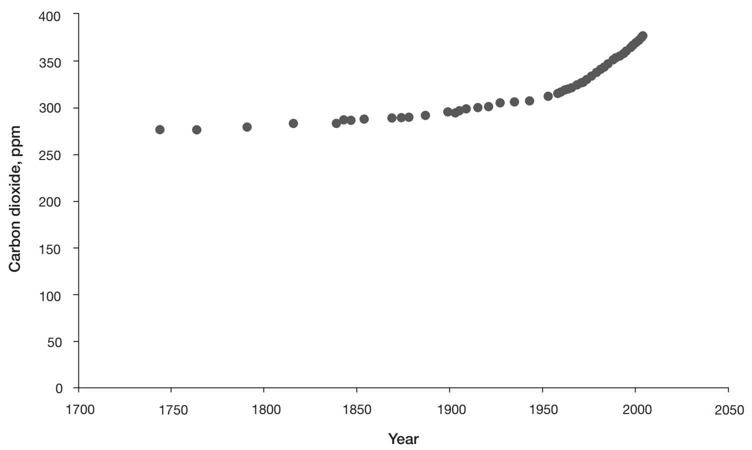


Fig. 6.1 The rising level of atmospheric CO_2. The graph shows the increase in atmospheric CO_2 concentration from its pre-industrial level of ~277 ppm in the middle of the 18th century to ~378 ppm in the first decade of the 21st century. Source: Values for 1744–1955 obtained by analysis of atmospheric gases in Antarctic ice cores: Neftel et al. (1994) in *Trends: A Compendium of Data on Global Change*. CDIAC, Oak Ridge National Laboratory, US Dept of Energy, Oak Ridge, Tenn. Values for 1958–99 obtained from atmospheric samples at Mauna Loa, Hawaii: Keeling, C.D. & Whorf, T.P./Carbon Dioxide Research group, Scripps Institute of Oceanography, La Jolla, Calif.

For any given level of global warming, science tells us the likely associated consequences. Not just a simple increase in temperature, but a rise in sea level with flooding of coastal cities and low-lying agricultural land, a spread of insect-borne diseases, movement of rain-bearing weather patterns away from their normal locations, an increased frequency of catastrophic weather events and many more.

When it comes to minimizing the impact of global warming, once again we must turn to science. To reduce our rate of emission of CO_2 into the atmosphere we must seek to replace fossil fuels with renewable energy sources, and this will require a range of science-based technologies. For electrical energy generation we need methods which derive their energy ultimately from the Sun. As pointed out in a recent US Department of Energy report[15], more energy from sunlight strikes the Earth in one hour (4.3×10^{20} joules) than all the energy consumed on the planet in a year (4.1×10^{20} joules). This effectively unlimited power supply can be harvested by a variety of methods. These will include efficient and cheaply-produced silicon solar panels, wind turbines and possibly ocean thermal energy conversion, which makes use of the ~20°C temperature differential between the surface and deep water in the tropical oceans. For vehicles, we will need liquid biofuels derived from photosynthetically produced biomass. For spark-ignition engines we already have ethanol which can be derived from sugar or starch, and for diesel (compression-ignition) engines we have vegetable oil, or preferably the methyl esters derived from the fatty acids in the oil. There are two problems here which must be taken into account. The first is that we do not want to use good farm land for biofuel production where it is needed for food crops. The second is that we must always pay close attention to how much liquid fuel is required to grow our biofuel crop. How many litres of diesel fuel does the tractor consume to produce 1000 litres of, say, canola oil. The net gain is often not very great. For these reasons the prospects for a substantial contribution to our liquid biofuel requirements on the basis of starch and vegetable oil crops grown on arable land are not good. There will be local exceptions,

especially in warm, wet tropical and sub-tropical regions. Sugar cane, being what is known as a C_4 grass species, uses both sunlight and water highly efficiently and is well suited for production of ethanol by fermentation of the sucrose it produces in high yield. This is already a major industry in Brazil, which gets a quarter of its transport fuel in the form of sugar cane ethanol. For significant production of biodiesel, palm oil is a potential candidate: since oil palms are perennial plants, less input energy should be required per ton of oil than in the case of annual vegetable crops such as canola or sunflower.

Looking to the long term, a possibility with enormous potential is the production of liquid fuel from total plant biomass, whether wood from forests or agricultural wastes – sugar cane bagasse, corn stover etc. Carbohydrate energy is there in the form of cellulose and hemicellulose. The problem is that these polysaccharides in the plant are, for very good biological reasons, complexed with lignin and this makes it impossible to simply hydrolyse them to sugars with enzymes. This, however, is a soluble problem on which research is already underway[16]. Heat and dilute acid, for example, can be used to loosen up the bonding and render the material susceptible to enzymic hydrolysis. High temperature thermochemical methods can be used directly on the biomass to produce volatile organic molecules for conversion to liquid biodiesel, and also hydrogen. There is much research still to be done, but the prospects are good. One of the great advantages of total biomass is that relatively little energy is required for its production. Suitable tree species, for example, can be coppiced. The trees are cut every few years: new shoots grow from the base which can then be harvested again. We can envisage the planting of energy forests for this purpose. The US Secretary of Energy announced in 2006 the goal of replacing 30% of gasoline with ethanol by 2030, and this will have to come predominantly from biomass.

But we can do better than this. Many plant species already make hydrocarbons, for example the terpenoid compounds in pine resin or eucalyptus oil. There is no reason why we should not breed plants which

accumulate hydrocarbons of a composition precisely tailored to our needs, but to achieve this we shall undoubtedly have to use the ultimate in science-based plant breeding methods, that *bête noir* of environmental extremists – *genetic engineering*.

Apart from solar energy, the only other major energy source which does not increase atmospheric CO_2 levels is nuclear power. Although this is a major energy source for electricity generation in many countries – 15% in the USA and 75% in France – it has fallen out of favour in the Western world in recent decades because of concern about the safe disposal of long-lived radioactive wastes, and concern with accidents such as that at Chernobyl, and the possibility of terrorists gaining access to nuclear material. These concerns are well based, but global warming now seems so inevitable, and its consequences so dire, that the time has come to look again at the nuclear option. One technological option that would greatly reduce the long-term waste problem would be the development of fast-neutron, liquid metal-cooled nuclear reactors[6,9], in contrast to the existing slow-neutron, water-cooled systems. Fast-neutron reactors can consume their own waste, as well as the waste generated by other nuclear reactors. In addition they can consume the stable uranium-238 isotope (the dominant form in natural uranium), not just the two minor radioactive isotopes, and so make much better use of a scarce, non-renewable, resource.

Nuclear energy generation in its existing forms involves *fission*, the disintegration of very large unstable atoms into smaller atoms, accompanied by the liberation of energy. In the very long run, nuclear *fusion* may provide the energy we need with relatively little production of radioactive waste products. The mechanism is similar to that which liberates energy within stars. In the Sun, at temperatures of ~15 million degrees, hydrogen nuclei (protons) are forced together to create helium nuclei. This process is accompanied by a small reduction in mass, with a corresponding release of energy. The method being explored for terrestrial nuclear fusion involves deuterium and tritium, two heavy isotopes of hydrogen. When the nuclei are forced together at temperatures in the region of 100 million K they

fuse and then break apart again to give rise to a helium nucleus and a neutron. There is an overall reduction in mass and a release of energy. The technical problems involved in this kind of energy generation are very great, and not likely to be solved any time soon. Research nevertheless continues, especially at the Joint European Torus facility in the UK.

Destruction of the soil resource

To grow food and fibre, wood for construction or fuel, and plant-derived industrial feedstocks, we need to maintain the soil inventory of the planet in a productive state. Apart from the obvious problem of agricultural land being expropriated for urban expansion, the main threats to soil productivity are *erosion* and *salinization* due to mismanagement of agricultural ecosystems. These are ancient problems for which science cannot be held responsible. Indeed, they preceded science by millennia.

Consider erosion first. This is the loss of soil by the action of water and wind on the unprotected soil surface (Fig. 6.2). Erosion may remove just

Fig. 6.2 Soil erosion caused by water flow in agricultural land.
Source: Courtesy of USDA Natural Resources Conservation Service.

the topsoil – exposing the less fertile, and usually texturally unfavourable, subsoil below – or may proceed right down to the bedrock. Soil takes thousands of years to form by the slow weathering of rock, so to lose it in our generation is for all practical purposes to lose it forever. Many of the bare rocky slopes around the Mediterranean were once covered with a mixed forest of oaks, beech, pines and cedars.[12] Following the deforestation that occurred in classical times, most of the soil was lost. The characteristic *maquis* plant complex that survives today, with its shrubs and aromatic herbs, is often regarded as the quintessential Mediterranean vegetation. It is in fact a degraded ecosystem, the bare remnant of something that was once much richer and more productive.

Land in the moist tropics is particularly susceptible to erosion because of the frequently torrential rain. Colonization of the islands of St Helena and Mauritius in the 17th and 18th centuries was followed by very serious soil erosion resulting from removal of the forests, causing great concern to the colonial administrations of the time.[5] Erosion continues apace in the tropics today. In poverty-stricken Pacific islands, timber rights are issued indiscriminately to unscrupulous Asian logging companies, forests on steep slopes are clear-felled and precious irreplaceable topsoil washes off into the sea.

While the main causes of soil erosion have been known to observant farmers for centuries – removing from the higher ground those forests which act as a sponge to absorb and slow down water movement, ploughing soils for crops on steep slopes, overgrazing with livestock so that all vegetative cover is removed – science can now offer ways of avoiding or minimizing the problem. Soil scientists can classify different soil types with respect to their inherent erodibility, and advise farmers as to the maximum slope on which cultivation can safely be carried out. On some soils *minimum tillage* farming, in which the existing weed or pasture growth is killed with a herbicide before a crop is planted, works well and avoids the necessity to stir up the soil by ploughing and other cultivation activities, thus lessening the erosion risk. In horticultural industries such as grape-

and fruit-growing, it is now common to control weeds in the immediate vicinity of the vines or fruit trees with herbicide sprays, thus avoiding the former standard practice of regularly cultivating the vineyard or orchard soil, with all the resulting damage to soil structure and danger of erosion. These various environmentally advantageous uses of biodegradable and short-lived herbicides have only been made possible by science – in this case organic chemistry.

Looking to the future, another way in which science may one day help is the development of perennial forms of our major crop plants. All our existing cultivars of cereals such as wheat, barley, corn and rice; pulses such as peas and beans; and oilseeds such as sunflower and canola are annual plants. That is, within the space of one year they are sown, they germinate and grow to maturity, are harvested and then die. But there is no fundamental reason why they should not have a perennial growth habit, i.e. after harvest the rooted base of the plant would survive, go dormant in the winter, start growing again in the spring and produce another crop. Every gardener is familiar with perennial plants which behave just like this. If our major crop plants could be made perennial they would not have to be sown every year. If they had to be resown only at intervals of several years there would be a many-fold reduction in the amount of soil cultivation required, and therefore a proportionate reduction in the extent of soil erosion. To introduce perenniality into crop plants will be very difficult (although some studies have in fact commenced), and the best hope of achieving it requires moving genes from unrelated plants – genetic engineering.

The second major threat to the world's soil resource is salinization. This is the phenomenon in which soluble salts, mainly sodium chloride, accumulate in soil and eventually reach a level which inhibits plant growth. This is most commonly associated with irrigation, and all over the world the productivity of irrigated agriculture is under serious threat from salinity. With regular application and evaporation of slightly salty irrigation water over a long period of time, salt concentration in the groundwater can increase. If the groundwater is near the surface it can rise by capillary

action then evaporate to deposit visible crystalline salt. In some places, notably in parts of Australia, the groundwater is naturally salty anyway so when the water table (the upper surface of the layer of groundwater in the soil profile) rises due to excessive irrigation it can reach the surface and, as above, be concentrated to toxic levels by evaporation.

In places where the groundwater is naturally salty we can have the phenomenon of *dryland salinity*. This is caused by removal of forest, not by irrigation. Trees, being perennial and deep-rooted, can use their continual transpiration of water through their leaves to transfer water from the soil to the atmosphere and in this way hold the water table well below the surface. When the forest is removed so that the land can be used for crops or pasture, the water table rises and the saline groundwater is concentrated by evaporation at the surface to the point that plant growth becomes impossible. Very large areas of non-irrigated agricultural land in Australia are now useless for production for this reason and a great deal more is destined for the same fate.

Salinization of agricultural systems is as old a problem as erosion. The earliest substantial agricultural civilization – that of Sumer in Mesopotamia, using water from the Tigris and Euphrates rivers to irrigate crops of wheat and barley – eventually petered out as agricultural productivity declined due to salinization of the fields.[12]

If we now understand salinization and have some idea of what can be done to avoid it, and in certain cases to reverse it, it is because of the research carried out over the years by soil scientists and hydrologists (scientists who study the movement of water in the landscape). The key is good drainage. If effective drainage pipes and channels can be put in, the system can be periodically flushed through with enough water to dissolve the accumulated salt and carry it into the river (leading to an unavoidable but temporary impairment of river water quality), and from there to the sea from which the salt came in the first place. Locations with soil which cannot be effectively drained are unsuitable for the establishment of irrigation agriculture, so accurate diagnosis of soil type is crucially important before any such development goes ahead. Dryland salinity is unfortunately

very difficult to reverse, and here the scientists' recommendation – alas, frequently ignored – is that further forest clearance should not be carried out where it is likely to exacerbate existing salinity problems.

LOSS OF BIODIVERSITY AND BIOLOGICAL RESOURCES

There are two environmental problems which need to be distinguished, although they are closely related. The first is the progressive and irreversible loss of biological diversity in the living world – the extinction of animal and plant species. For the purpose of this chapter we shall simply assume that this is inherently a bad thing, although to establish that proposition is not at all easy, as we discuss in Chapter 8. Let us for the moment adopt the ethical position that – to quote the eminent zoologist Sir Robert May – we should leave a world as rich as the one we inherited. The second environmental problem is the loss – either through complete extinction or greatly reduced availability – of biological species or whole ecosystems, that we human beings value for purely practical reasons such as that they provide us with food, timber, useful chemicals or simply recreation. Thus we shall distinguish between the loss of biodiversity and the loss of biological resources, although the two often go together.

That we have within historical times lost, and continue in the present day at an ever-increasing rate to lose, large numbers of animal and plant species is now well known. Accounts of this unfolding catastrophe can be found in the writings of experts such as Edward Wilson.[17] It has always been the case since life began that species have become extinct as new species have come into existence. But once *Homo sapiens* appeared on the scene about 200 000 years ago, with mental capacities vastly superior to those of any other species, the rate of extinction began to increase. At first it was large herbivorous prey species that fell victim – especially as *H. sapiens* spread out into the new lands of the Pacific[3] where such species, never previously having encountered humans, did not recognize us for the

dangerous predators, we are. But when human beings increasingly began to modify the environment – at first by practices such as burning the forest to maintain an open structure better for game (the 'firestick farming' practised by Australian Aborigines) and then by agriculture – the rate of extinction began to increase until species are now disappearing enormously faster than new ones are appearing.

While we know plenty of examples of extinctions in historical times it is very difficult to put a figure on just how fast species are disappearing now. We know that the giant bird, the dodo, was exterminated on the island of Mauritius in the 17th century, and that another giant bird, the moa of New Zealand, was hunted to extinction by the Maori within a few centuries of their arrival in that land 800–1000 years ago. We know that more than half the bird species present in the island of Hawaii, before contact with the western world, are now extinct, and many of us have seen the sad film footage of the last surviving Tasmanian tiger (a wolf-like marsupial carnivore – Fig. 6.3) as it paced to and fro in its cage in the 1930s. The passen-

Fig. 6.3 The now-extinct Tasmanian tiger (*Thylacinus cynocephalus*). It was a wolf-like carnivorous marsupial native to Tasmania (Australia). It had light-brown fur with dark stripes across its lower back, hence the name 'tiger'. This 1933 photograph shows the last Tasmanian tiger in captivity, in Hobart Zoo.
Source: Courtesy Tasmanian Museum and Art Gallery.

ger pigeon of North America, which was once so numerous that its flocks darkened the skies, was shot to extinction in the early years of the 20th century. Specific examples such as these make the phenomenon of species extinction real, but it is hard to proceed to actual numbers.

One of the main reasons for this is that we still do not know how many species there are on this planet. It is not unexpected that this is true of unglamorous creatures such as the tiny nematodes, mites and springtails etc. that make up the soil invertebrate fauna. What is more surprising is that it is also true of higher, more attractive, organisms such as mammals and birds. Since 1990, 10 new species of monkey have been described, and new bird species are still being found at the rate of about 3 per year. The number of as-yet-undescribed insect species – butterflies, beetles, grasshoppers, moths, mantises, dragonflies etc. – could be in the millions. Estimates of the total number of species existing on this planet vary wildly but a commonly expressed order-of-magnitude figure is that there are in the region of 10 million species, of which only about 1.4 million have so far been described. Even though we cannot give a definite figure, there are so many well-documented examples of species that have ceased to exist within historical times, many of them species which we would dearly like to have back again, that the reality and seriousness of the problem cannot be doubted.

The second of our two closely related environmental problems is loss of biological resources. We are mainly talking here about animal and plant species in natural ecosystems, which human beings have traditionally harvested for their own use. In these cases the most common danger is not the total extinction of the species, but depletion of the numbers to the point that for all practical purposes it is no longer available for our use. The world's fisheries provide innumerable examples of where populations of particular highly regarded species have collapsed to the point that fishing had to be curtailed to allow recovery: even plentiful species such as the Atlantic cod and the North Sea herring have not been immune. In the plant kingdom, many desirable forest timbers are now no longer available. When the early colonists settled New South Wales, Australia, in the 19th

century, they made extensive use of a particularly high-quality native tree, the so-called Australian cedar *Toona australis*, which was then quite plentiful. It is virtually unobtainable today.

A particular worry is the loss of biological resources, especially from the tropical rainforest, that we don't even know we have. Plants are marvellous organic chemists and make a vast array of different compounds. Many of these are already known to have valuable pharmacological activity and indeed are used for this purpose. One good recent example is the potent anti-cancer drug *taxol*, which is extracted from the bark of the Pacific yew tree (*Taxus brevifolia*). It has a complex and peculiar molecular structure (Fig. 6.4), and although it can be made in the laboratory it is very unlikely

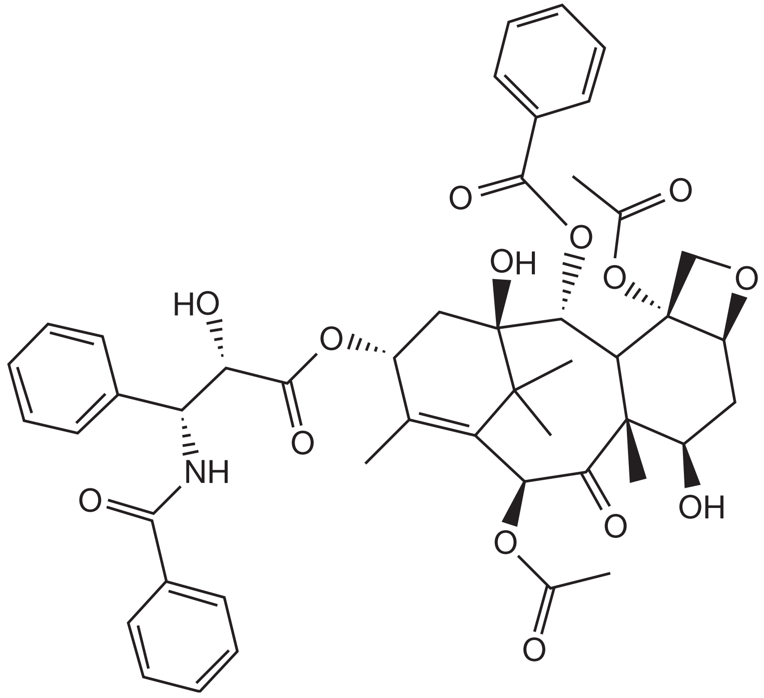


Fig. 6.4 The complex molecular structure of the naturally-occurring anti-cancer drug taxol. This was isolated from the bark of the Pacific yew tree (*Taxus brevifolia*). Because of the drug's importance, a chemical synthesis has been achieved, but it is very unlikely that any organic chemist would otherwise have bothered to construct this bizarre molecule.

that it would have occurred to any human organic chemist to synthesize it. We can confidently assume that in the hundreds of thousands of plant species in the tropical rainforests and other ecosystems, there are valuable biochemicals waiting to be discovered. If the species go extinct because of indiscriminate forest clearing, these irreplaceable resources will be gone forever. We can also assume that many, as yet uncharacterized, tree species in such forests produce valuable wood. If these were recognized in time they could be propagated for timber production in plantations.

What are the main causes today for the continuing loss of biodiversity and biological resources? The most important is destruction of habitat. Every living species has a particular niche in a particular ecosystem. If that ecosystem is destroyed, then in the vast majority of cases the species cannot survive other than by being artificially kept alive in zoos or botanic gardens. There are of course a few 'weed' species such as thistles, rats, cockroaches and starlings which thrive in many kinds of environment, but most species are not so adaptable. Take away the complex web of ecological relationships in which they are normally embedded, and their numbers quickly decline. Everywhere on the globe, as the human population has expanded and more and more land has been taken over for agriculture and urban expansion, important ecosystems have been greatly truncated or have disappeared altogether. And the process continues, especially in the third world where a burgeoning population presses continually up against a limited resource base.

Introduction of alien species to new environments has frequently had disastrous consequences for biodiversity. The European rabbit, introduced as a game species to Australia in the mid-19th century, found an environment to which it was very well suited and it proliferated enormously. The resulting massive rabbit population overgrazed and destroyed much of the native vegetation, leading to loss of species as well as serious soil erosion problems. Making matters worse, the red fox was introduced to control the rabbit population. Not only did it fail but it began immediately to prey upon small native marsupial species, some of whom – never having seen

foxes before – were slow to recognize them as predators. Thus, animal biodiversity was also diminished. The small native animal species suffered further depredation from feral populations of another introduced species – the domestic cat. Introduced plant species can also be disastrous as they spread without check in the new environment which lacks the diseases and insects which limit their proliferation in their native habitat. A typical example is the rubber vine *Cryptostegia grandiflora*, a native of Madagascar, which was introduced to Australia in the 1870s and has become a very serious weed in the tropical north of Queensland. It currently infests about 700 000 hectares along the river systems. During a sojourn with scientific colleagues from the University of California, Santa Barbara, at the field station on Santa Cruz Island off the Southern California coast in 1989 I was able to observe how the introduced European plant, fennel (*Foeniculum vulgare*), presumably brought in as a herb, had come to dominate the low-lying land and creek beds throughout the island. As these alien introduced weed species proliferate they can eliminate native species by outcompeting them. A particularly telling example of the effect of alien species on biodiversity is the island of Hawaii where most of the existing species of plants and animals are introduced – native species have largely disappeared.

In the case of our second environmental problem, the depletion of biological resources, overharvesting is the main cause. Experience shows that if the management of any fishery is left to the fishermen engaged therein, they will without fail fish it to the verge of extinction. Most of the world's commercial fish species are now fished at a rate beyond, or just about at, that which is sustainable. Some highly prized and once plentiful species are now hard to obtain. To give one example, the number of adult bluefin tuna in the Western Atlantic is estimated to be only about 10% of its 1970 level. The timber resources of forests are also routinely over-exploited. We have already mentioned the example of Australian cedar. The famous cedar forests of Lebanon (no relation) were decimated even in biblical times. A disastrous secondary consequence of over-exploitation of forests is that if it is carried out by clear-cutting on steep slopes there is

very likely to be serious soil erosion, with a consequent diminution (or in the worst case complete abolition), of the ecosystem's ability to regrow. Clear-cut harvesting of native forests for timber also has the potential to cause inadvertent extinction of non-target (not used for timber) plant species, as well as of the animal species which depend upon them.

An additional worrying problem in relation to forest clearance, on top of the biodiversity and biological resources issues is – what effect would massive clearance of tropical rainforest, especially in South America, have on the global hydrological cycle? At the moment these warm evergreen forests continually transpire enormous amounts of water into the atmosphere. This is subsequently returned as rain. If the forest is cleared and replaced by low-quality, shallow-rooted grazing land, will movement of water through the system continue in the same way? If not, the effects could be felt well beyond the cleared land itself.

Pollution of various kinds can also be a threat to biodiversity and biological resources. In less enlightened times, sulphur dioxide in exhaust gases from smelter chimneys in metal-mining areas in Canada and Australia devastated the vegetation for miles around. Around Queenstown, Tasmania, there are hilltops consisting only of bare rock, their original tree cover having been killed by smelter SO_2. While unsightly on a local scale, however, this is not a major problem globally. The pollution of the atmosphere by fossil fuel-derived CO_2, on the other hand, may well have effects on biodiversity and biological resources on a global scale. It is not the CO_2 itself which is a problem: indeed, the increased CO_2 levels should lead to increased plant growth. Rather it is the general warming that seems likely to accompany the increased CO_2 that is the difficulty. Temperature is a key environmental variable to which ecosystems are very sensitive. An ecosystem well adapted to an average annual temperature of, say, 15°C is likely to be ill-adapted to one of 17°C or 19°C. In principle, with global warming we might envisage ecosystems shifting to higher latitudes where they will find the temperature regime they need. In reality, although ecosystem movement can occur it is very slow – much slower than the

predicted rate of global warming. It seems more than likely that the ecosystem dysfunction induced by temperature change will lead to substantial loss of species. Some ecosystems cannot move, namely those which occur only on certain islands. The unique cold-adapted flora of Macquarie Island, for example, at 55°S in the Southern Ocean, has no nearby landmass and will have nowhere to go if global warming occurs.

Pollution of the aquatic environment tends on the whole to be more serious in its biological effects than pollution of the terrestrial environment. Discharge into rivers and estuaries of effluent containing oxidizable organic matter can lead to the rapid depletion of oxygen in the water as bacteria use it to consume the organic compounds, with the consequent death of all aerobic organisms, such as fish and crustaceans. Surface drainage runoff containing nitrogen and phosphorus from farmland, or treated sewage effluent which still contains these nutrients, can trigger massive overgrowth of green or cyanobacterial algae which can smother and kill the complex array of other algae and vascular aquatic plants, with their associated animal life-forms, which previously existed there. Coral reefs are particularly susceptible to being overgrown by green algae as a consequence of nutrient pollution. Dredging for sand or shell, or even to maintain navigation channels, within estuaries causes a great increase in water turbidity – optical, rather than chemical, pollution – and this is particularly disastrous for seagrasses, plants that are especially sensitive to any diminution in ambient light levels. Since seagrass beds are of great ecological and fishery importance, providing nursery areas for young fish as well being home to a wide range of invertebrates, damage to them is likely to have negative implications for both biodiversity and biological resources.

The rise in ocean temperatures in response to increasing atmospheric CO_2, which already seems to be detectable, will certainly have far-reaching biological effects. The consequences are likely to be particularly serious for coral reefs. It seems to be commonly the case, and is already being observed, that even when a comparatively modest rise in water tempera-

ture occurs corals lose their symbiotic algae, the *zooxanthellae*, whose photosynthesis is their main source of energy. We can only hope that there are other coral species waiting in the wings, perhaps currently minor components of the system, which will be better able to cope with the higher temperatures.

What has science to do with all this, with the loss of biodiversity and the depletion of our biological resources? Clearly it has everything to do with it. Science is not responsible for these environmental problems. It did not cause them. But it is our only hope of understanding them and coming up with solutions. The very recognition of the existence of such a thing as biodiversity, and the characterization of the innumerable species, genera, families etc. of which it is constituted, was the product of biological science. It was professional botanists, zoologists and ecologists who raised the alarm about the accelerating loss of ecosystems and their associated life-forms. These points are so obvious that I will not labour them further. Let us rather proceed to a consideration of possible solutions to these problems, and the essential role that science must play in finding them.

To slow down the loss of biodiversity, the crucial element is the conservation of habitat. While zoos and botanic gardens can have a role in an emergency, without doubt wild species survive best in their own ecosystems. Thus, representative samples of all major ecosystems need to be conserved, and we now know that these have to be large. Small islands of ecosystem surrounded by modified agricultural/industrial landscape will not work. Populations – particularly of species higher up the food chain – need to be above a certain minimum size or they go inexorably into decline. It is to the biological scientists that we must turn, as we already do when we wish to establish national parks, to identify the key ecosystems and advise on the selection and maintenance of the conservation areas. Where, as in the case of the Amazonian rainforest and certain other tropical forests, we simply do not yet know what biodiversity and biological resources exist within them, the only prudent and scientific policy is to keep the removal of such forests to an absolute minimum. Their possible key role in the global hydrological cycle is another reason for prudence.

We cannot of course deny underdeveloped countries with large areas of forest their right to use some of it, as Europeans did in their continent centuries ago, for agriculture and timber production. What science can do is advise what specific areas are best suited for conversion to agriculture and forestry, by what methods these activities should be carried out so that they are sustainable in the long run, and what areas of forest should be set aside for conservation of biodiversity and biological resources.

Turning to that other threat to biodiversity, namely the invasion by alien species, undoubtedly the best solution is biological control. The main reason why a certain species of plant or animal can sometimes proliferate massively when introduced to a new environment is that it is free of the natural enemies which held its population in check in its original home. The way to control it, then, is to identify those enemies in the original habitat and transfer one or more to the new environment in the hope that they will achieve effective control. Stringent tests of course have to be carried out to ensure that the transplanted control organisms will be highly specific in their target and not start attacking blameless native species in their new home.

There are many success stories in the field of biological control. In the early years of the 20th century, *Opuntia* cactus (a US native) became rampant through much of the hotter and drier parts of Australia. After introduction of the *Cactoblastis* insect, which feeds on *Opuntia* in its native home, the cactus population collapsed and has not been a serious problem since. To control the rabbit population two viruses, first myxoma and more recently calicivirus, spread by mosquitoes and other insects, were introduced to Australia. While they have not eliminated the rabbit in the way that *Cactoblastis* has virtually eliminated *Opuntia,* they have nevertheless made serious inroads into the rabbit population. The aquatic weed *Salvinia molesta,* originating in South America, which has the capacity to completely cover the surface of water bodies in warm regions, was very effectively controlled in inland waters in Australia and New Guinea when its natural enemy, the weevil *Cyrtobagous salviniae,* was introduced. In the

case of the rubber vine, a serious weed in Queensland, a possible biological control currently under trial is the fungus *Maravalia cryptostegiae*, which invades the leaves and causes them to drop.

Biological control, when it works, is not only amazingly effective but is also, from an environmental perspective, vastly superior to any other methods of controlling invasive pests. Typically, once the control agent has been established it maintains itself without any further expenditure. The alternatives, such as chemical spraying, are not only environmentally problematic but require expenditure and effort extending indefinitely into the future. But it goes without saying that to achieve biological control of an invasive alien species the services of appropriate professional scientists such as entomologists, plant and animal pathologists and ecologists are required.

As far as maintaining the world's biological resources is concerned, the role of science is particularly clear. The principles of management of fish stocks, for example, have been well established since the early 20th century. From the age distribution within a given population of fish – which of course requires regular sampling and analysis by competent professionals – it is possible to determine the rate of reproduction, and from this the *sustainable yield* (the rate of harvesting which will permit that population to survive at a usefully high level for the indefinite future). If analysis shows that a population is composed of, on average, very old individuals this means that even though the population may be large its reproduction rate is slow, so the sustainable yield will be a correspondingly small fraction of the total existing biomass. Such is the case with a fish relatively new to commerce, the *orange roughy* (*Hoplostethus atlanticus*) (Fig. 6.5), large populations of which have been found in deep water around certain seamounts in the Southern Ocean. This at first looked like a plentiful new resource, but analysis of the age structure of the population showed that these fishes were in fact extraordinarily old, the average age being in the region of 50–100 years. This means that the sustainable yield of orange roughy, despite its apparently large population, is in fact quite low. Sadly,

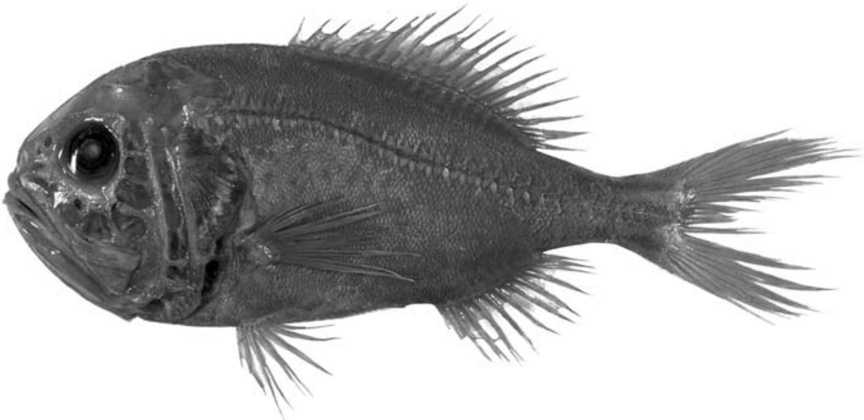


Fig. 6.5 Orange roughy.
Source: Courtesy CSIRO Marine and Atmospheric Research.

while civilized governments in the southern hemisphere are endeavouring to manage the fishery in a sustainable way, rogue trawlers from certain other nations are plundering the resource without regard for its long-term survival. This reminds us that science alone is not enough to maintain our biological resources – informed and resolute government action is also required.

In the case of fisheries, marine scientists say that governments can help by creating marine reserves where removal of fish or other marine resources is banned. Studies of 81 marine reserves around the world have shown dramatic increases in population densities, organism size and species diversity within these areas.[14] Provided such reserves are big enough, we can be cautiously confident that even species which are heavily exploited outside the reserves will survive within them. Furthermore, when target species are fished to near-extinction they will have a good chance of rebuilding from the reserve populations, provided commercial fishing for the species is halted for a sufficient time.

Even if every fishery in the world is properly managed, however, it is now quite clear that there are not enough fish to go round. The present

demand, let alone that of an increased population in the future, exceeds the sustainable supply from all the world's oceans. If we want more fish we must turn again to science, specifically to aquaculture, a technology which is heavily dependent on science as it relies on a battery of sophisticated procedures such as hormonal control of breeding cycles, precise specification of feedstuff composition, monitoring of water quality, control of pathogenic micro-organisms and so on. Although aquaculture has a great potential for increasing global fish and crustacean production, before we give it the green light for unlimited expansion we must realize that it also has a great potential for causing environmental problems of its own. Three of these are particularly serious. The first is that aquaculture actually consumes very large amounts of fish – of inferior kinds, but still harvested from the sea – to feed the fish or shrimps being farmed. What is needed is replacement of fish protein with high-quality plant protein of the right amino acid composition; some studies on this using genetic engineering have already been carried out. The second major problem is that fish/crustacean farms are very often established on what were coastal wetlands, thus destroying valuable and increasingly scarce aquatic ecosystems. Ironically, these are important nursery areas for many of the fish populations in the commercial fisheries of the sea. The third problem is that, like all intensive animal rearing, aquaculture unavoidably produces large amounts of effluent which has the potential to cause serious pollution problems in any waterway to which it is discharged. Effective effluent treatment technology has to be part of the whole package.

Management of the world's forest resources is, like that of fishery resources, a mature science-based technology. Professional foresters know in general terms how to manage a forest for timber supply on a sustainable basis, and are well aware of the need to avoid problems such as erosion caused by clear-cutting large areas on steep slopes. Where we see forests being seriously mismanaged, as in parts of the Pacific, it is because foresters' advice is not being taken. Plantations of selected species are a very effective method of maintaining the availability of valuable tree

resources, and forest scientists are needed for the selection and propagation of the best genetic material. In some cases, forests need to be managed for the production of medicinal plants as well. To ensure that such plants are not lost, the ideal method would be for natural products chemists to identify the best individuals in the population then get them out into cultivation.

BACTERIAL RESISTANCE TO ANTIBIOTICS

Everyone who reads the papers or listens to the news via the electronic media must be aware of the looming problem of antibiotics' diminishing effectiveness against pathogenic bacteria. When antibiotics, particularly penicillin, first became readily available for general medical use after the Second World War they were marvellously effective. They were true 'magic bullets' which hit the mark and eliminated the pathogen – if it was a gram-positive bacterium such as *Staphylococcus* or *Streptococcus* – practically every time. But before very long occasional strains of bacteria resistant to the drugs were observed, and the incidence steadily increased so that in the last two decades antibiotic resistance has become common, especially in the widespread pathogen *Staphylococcus aureus.* As bacteria became resistant to one kind of antibiotic other kinds were tried and initially found to be effective, then resistance to the newer kinds of antibiotic also emerged. We are faced now with the situation where the medical profession has only one shot in its locker about which it can usually be confident, namely the antibiotic *vancomycin.* Even this is proving to be not totally reliable, since some vancomycin-resistant strains of *Staph. aureus* have now been detected. Furthermore, *Mycobacterium tuberculosis,* the causative agent of the frequently fatal lung disease tuberculosis, is now generally resistant to the antibiotic *streptomycin* which was for many years a very effective treatment for the disease.

While the catastrophe of generally uncontrollable bacterial infection is not yet upon us, it is looming on the horizon. Although this is, of course, a

medical problem, we must also regard the fact that the human environment is now widely contaminated with antibiotic-resistant bacteria as an environmental problem, just as much as if it were contaminated with toxic chemicals. Since antibiotics are very much a scientific phenomenon, involving microbiology, biochemistry and pathology, this problem can only be understood and addressed in terms of science. But science did not cause this problem: rather, the misuse of science's great gift to human health is at the heart of the matter.

How does bacterial resistance to antibiotics arise? In the first instance, by mutation. In rapidly multiplying organisms such as bacteria there are always mutations going on, in which there are localized changes in the base sequence of the DNA caused by copying errors, high-energy particles from radioactive decay or cosmic rays, ultraviolet light and mutagenic chemicals. The vast majority of such mutations are either neutral (have no effect) or are harmful to the organism. But very occasionally a change will take place which is useful to the organism, or at least useful in some specialized environment. Resistance to an antibiotic to which an organism was previously sensitive can arise in a number of ways, depending on the antibiotic's mode of action. At its simplest, if a particular antibiotic begins its inhibitory action by binding to a certain protein in the bacterial cell, a mutation which changes the structure of that protein in such a way that the antibiotic molecule can no longer bind will confer resistance. Thus, in an environment containing the antibiotic, such as the gut of a person or an animal being treated with the drug, the mutated bacteria will have an advantage and by being less inhibited than the rest of the bacterial flora will multiply to become a major component of the population.

Quite apart from the acquisition of antibiotic resistance in the target population, bacteria can acquire resistance genes directly from other bacteria which already have them. There are three known mechanisms for this. In *conjugation,* two bacterial cells are temporarily joined by a narrow tube along which DNA passes from one to the other. In *transduction,* a virus known as a *bacteriophage* grows inside one bacterial cell and incorporates

some of the host DNA into its own DNA. When virus particles are released from the first cell and one of them infects another, this piece of DNA can be transferred to the DNA of the new host. Conjugation and transduction take place between cells of the same bacterial species. In *transformation,* the third mechanism, bacteria simply take up DNA which they encounter in the environment after it has been released by other bacteria; in this case bacterial genes can cross the species barrier. Given the existence of these three mechanisms for gene transfer, it is easy to see how rapidly antibiotic resistance can spread in the microbial world. Indeed, the mechanisms allow harmless bacteria, which are not themselves pathogenic but which happen to have antibiotic resistance genes, to transfer resistance to dangerous pathogens.

In the light of our understanding of what is happening at the gene and cell level, can we identify what aspects of antibiotic usage may have contributed to the rapid appearance of resistance? Looking to the future, can we do something to stop things getting even worse? There is now general agreement in the medical microbiology community that the main culprit is overprescription. Doctors – often in response to direct patient pressure – will frequently prescribe antibiotics when they are not essential, and even sometimes in cases such as viral infections where the drug will do no good. The problem is exacerbated when patients do not complete their course of treatment. To minimize the likelihood of resistance appearing, it is best to totally eliminate the target bacterial population. To cease antibiotic usage when the population has been extensively exposed to the drug but has not quite been wiped out is a strategy tailor-made for selecting resistant types. As the bacteria again begin to multiply in the presence of inhibitory but now declining and sublethal levels of the drug, resistance may emerge and be selected for.

A much less forgivable misuse of antibiotics is in agriculture, specifically in livestock rearing, and certain kinds of horticulture. As early as the 1940s it was found that continual feeding of low levels of antibiotics to intensively reared cattle, pigs and poultry increased their growth rate, and this

became standard practice in much of the developed world. In the US, more than 40% of antibiotic usage is for animals and of this about 80% is used not to treat sick animals but in animal feedstuffs. The practice of continually exposing the bacterial flora of very large numbers of animals to low levels of antibiotic might almost have been designed to generate antibiotic-resistant strains of bacteria and disseminate them into the environment. Even newly-developed antibiotics are being compromised in this way. Quoting from a recent editorial in Science[2] by Stanley Falkow and Donald Kennedy:

> *In 1996, the FDA [US Food and Drug Administration] approved the use of fluoroquinolines in chickens and turkeys … This inexplicable decision was reached despite strong opposition from the Center for Disease Control (CDC), which cited the extraordinary value of these compounds in treating community- or hospital-acquired enteric infections in humans. Subsequent events showed that the CDC's concerns were prescient. Fluoroquinoline resistance quickly appeared in Campylobacter isolated from chickens, and by 1999 17.6% of C. jejuni and 30% of C. coli isolated from human patients showed fluoroquinoline resistance. Campylobacter infections are the leading cause of food-borne illness in the United States.*

In the US, antibiotics are also used for bacterial disease prevention and control in commercial fruit production. Tetracycline and streptomycin are used for this purpose in apple and pear orchards. As these antibiotics are spread indiscriminately into the environment, bacterial populations will be exposed to low levels of the drugs with consequent selection in favour of resistant forms.

What can be done to slow, and if possible to reverse, the increasing incidence of antibiotic resistance among bacteria? Clearly, where possible we should cease activities which have been identified as primarily responsible for the spread of drug resistance. The message about overprescribing is being widely disseminated not merely through the medical fraternity but

through the population at large. Use of antibiotics as livestock feed supplements and in horticulture should cease forthwith. The intensive animal-rearing industries will no doubt issue dire warnings about consequent higher prices for their products. They may well be right but we can live with this, just as we did in the not-so-distant times before antibiotics were ever added to animal feedstuffs. We also need to take a cold hard look at using antibiotics for animals even for therapeutic purposes. Since human health is vastly more important than animal health, certain antibiotics should be quarantined entirely for use in human, not veterinary, medicine.

The search for new antibiotics, which apparently slackened somewhat in recent decades, should be renewed with vigour. The places to look will be not only among the traditional sources – moulds and soil micro-organisms such as actinomycetes – but also among the extraordinarily versatile chemists of the plant kingdom (another reason for conserving the rainforest) and the pharmacologically inventive invertebrates of the ocean. In addition to seeking those useful compounds which nature has already made for us, we may reasonably hope that, as we come to understand the functioning of bacterial cells at an ever-greater molecular level of detail, it will become possible to design completely new antibiotics precisely targeted at specific bacterial processes. Once again we will be looking to science, and science of a most sophisticated kind, to save our bacon.

POPULATION GROWTH AND OTHER ENVIRONMENTAL PROBLEMS

The account I have given is of the environmental problems that I believe to be the most serious. There are of course many other such problems, and other people would give some of them a higher priority than I have done. We shall examine them briefly here.

World population growth is the obvious example. This is undoubtedly a major problem, but my impression is that it is somewhat less threatening

than it once seemed. At the beginning of this new millennium the world population is about 6 billion people, but the rate of increase has apparently diminished. It is true that if the Earth's population grows beyond what the planet can support, it will be a disaster. But whether the population will rise to this extent remains uncertain.

An unanticipated problem which has crept up on us while our attention has been focussed on excessive population growth in the third world, is the collapse of birth rates in many developed countries. In some of these countries the birth rate is well below that required for maintenance of a stable population number. In the European Union, for example, the current fertility rate is around 1.5 births per woman[10], well below the replacement rate of 2.1.This is partly because women are delaying births to later ages when their fertility has already substantially declined. In the UK it has been found that only about half of women who postponed childbearing into their thirties, and who still intended to start a family, managed to do so in the subsequent six years.[11] Quite apart from the impending economic consequences – an ageing population, a progressive diminution of the proportion of the population able to do productive work – if birth rates substantially below replacement become a permanent feature of first world demography then some of the most creative nations in world history will cease to exist. It is as though certain nations have lost the will to live. Some centuries hence, will human beings look back on the Italians, for example (among whom this change is particularly evident) in the same way as we look back on the Etruscans: an interesting and creative but sadly now extinct people?

Other problems which I have not highlighted but which loom large in the consciousness of many environmentalists are radioactive contamination and chemical pollution of the environment. On a local scale these can be very serious, for example the radioactive contamination of a large area in the Ukraine after the nuclear explosion at Chernobyl, and the poisoning of thousands of people by the toxic chemical discharge at Bhopal in India. Other serious instances with very long-term consequences are the

discharges of toxic heavy-metal-containing mining wastes into waterways. A particularly inexcusable recent and ongoing example is the siltation and chemical pollution of the Fly River in New Guinea – one of the largest river systems in the world – by waste rock and copper-containing tailings from the giant Ok Tedi mine. This has led to the death of downstream forests and swamps, fish kills and loss of sago palms, an important local food source. Nevertheless, serious though such matters are, I did not include them in the main list because their effects are not global. Also, it is fair to say that these are problems of a type which we already know how to solve. Where they are not solved it is because of local incompetence, or a failure in social mechanisms.

A problem which is truly global but which I have nevertheless not chosen to treat as a major issue is depletion of the ozone layer. We will recapitulate the underlying science. The Earth's atmosphere consists of two major layers: the *troposphere*, extending from the surface up to about 10 km altitude, and the *stratosphere* at 10–50 km. Within the stratosphere the oxygen is exposed to the full glare of fierce solar radiation, including some highly energetic ultraviolet photons of wavelength less than 200 nm (0.2 microns). When an oxygen molecule absorbs such a photon, the energy causes it to dissociate into two oxygen atoms:

$$\mathrm{UV} < 200\ \mathrm{nm}$$

$$O_2 \rightarrow O + O$$

The single oxygen atoms then react with diatomic oxygen molecules to give the triatomic form of oxygen, ozone (O_3):

$$O + O_2 \rightarrow O_3$$

The ozone thus formed absorbs, as we saw earlier (Chapter 2), ultraviolet light of wavelengths in the 200–300 nm range, this being the mechanism by which the ozone layer protects us from solar UV. As the ozone molecules absorb this radiant energy they are in turn dissociated to give diatomic oxygen and single oxygen atoms:

$$\text{UV 200–300 nm}$$

$$O_3 \rightarrow O_2 + O$$

Some of the oxygen atoms produced in this or the earlier reaction can react with ozone to form diatomic oxygen:

$$O + O_3 \rightarrow 2O_2$$

and so the cycle is complete. All these processes are taking place simultaneously in the stratosphere, and the actual stratospheric ozone concentration is a steady-state level, determined by the relative rates of these reactions.

The depletion of the ozone layer – most dramatically manifested by the appearance in the last decades of the 20th century of a gigantic hole in the ozone layer above the Antarctic continent (Fig. 6.6) – is known to be mainly due to atoms of the halogen element, chlorine, in the stratosphere.

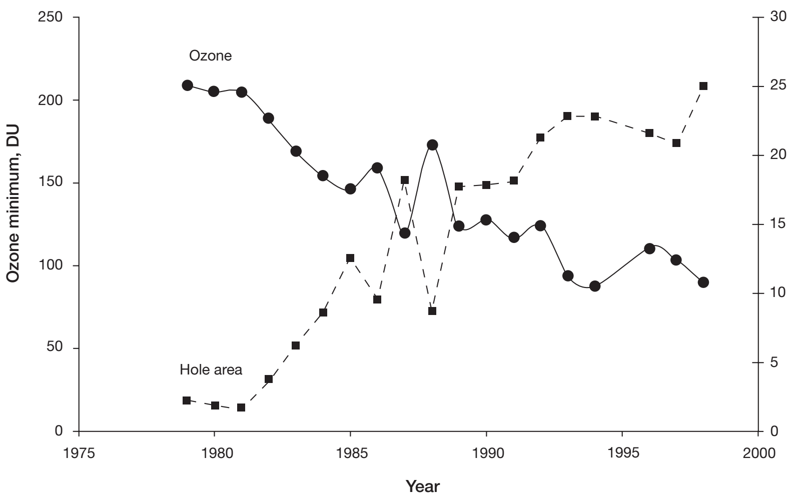


Fig. 6.6 Development of the hole in the ozone layer over the Antarctic continent. Since the early 1980s, a very large hole has developed in the ozone layer over Antarctica each Antarctic spring. This graph shows the minimum ozone concentration (•) in Dobson units (DU) observed each year, and the area (■) of the hole (in million km^2). Over most of the globe the ozone layer has a thickness of ~300 DU. At its largest, the hole has been bigger than North America.
Source: Data obtained from space by instruments on NASA satellites.

These have come predominantly from molecules of *chlorofluorocarbons* (CFCs) which were extensively used not only as the working fluid of refrigerator compressors, but also as aerosol propellants and for expanding plastics such as styrofoam. CFCs are chemically very stable here in the troposphere and are not washed out by rain, so they diffuse upwards, eventually reaching the stratosphere. They are no longer stable there, because they are exposed to intense UV radiation which dissociates them so that single chlorine atoms are liberated:

$$CF_2Cl_2 \xrightarrow{\text{UV} < 260\text{ nm}} CF_2Cl + Cl$$

The single chlorine atoms then react with ozone molecules to give diatomic oxygen and chlorine monoxide, ClO:

$$O_3 + Cl \rightarrow O_2 + ClO$$

The ClO then, with the help of UV light, reacts with another ozone molecule to give two molecules of diatomic oxygen, with the regeneration of the chlorine atom:

$$ClO + O_3 \xrightarrow{\text{UV}} Cl + 2O_2$$

This chlorine atom can go round the cycle many more times, each time breaking down another two molecules of ozone. Thus in chemical terms the chlorine is *catalysing* the breakdown of ozone, and it is calculated that on average one atom of chlorine can degrade over 100 000 molecules of ozone.

This, together with a lesser but real contribution from atoms of another halogen element, bromine (used in combined form – *halons* – in fire extinguishers, and as the fumigant methyl bromide), is the mechanism by which our protective ozone layer has been significantly depleted. The phenomenon occurs all around the globe although it is intensified at the poles, and the effects – melanoma, other skin cancers, cataracts, damage to materials and to some animals and plants – are serious. So why is it not one of the

major environmental problems we should be worrying about? The reason is that for this problem at least, matters are already in hand: the existence of the problem has been clearly recognized and the necessary remedial measures, specifically the banning of production of CFCs and halons, have been put in place. The ozone hole and the general thinning of the ozone layer will be with us for many years but the indications are that the problem has essentially been solved, and by about the middle of this century things will be back to normal.

Depletion of the ozone layer is an environmental problem for which the finger can fairly be pointed at Science. CFCs are not natural compounds. They were created by organic chemists and when their virtues – inertness, volatility, non-toxicity – were recognized, the compounds were enthusiastically taken up by the refrigeration, aerosol and plastics industries. But while science inadvertently (since little was known about the possible behaviour of chlorine in the stratosphere) created the problem, it was also scientists who discovered that the problem was developing. A research team in the British Antarctic Survey first observed a diminution in ozone over Antarctica in the 1970s, then in 1985 saw the catastrophic fall which we now call the 'ozone hole'. Again it was scientists – notably the atmospheric chemists Molina, Rowland and Crutzen, jointly awarded the Nobel Prize for Chemistry in 1995 – who determined the mechanisms responsible for ozone loss. And it was on the basis of all this good solid science that countries agreed to phase out production of CFCs and halons.

A comparatively recent scientific development, which has not yet caused demonstrable harm to the environment but about which concern is frequently expressed by some environmentalists, is *genetic engineering*. It has always been possible to manipulate the genetic composition of the organisms – plants, animals, microbes – that are useful to humanity. Plant and animal breeders have been doing it for thousands of years, since agriculture and domestication of animals began. The rate and effectiveness of such manipulation greatly increased when the principles of genetics were elucidated in the late 19th and early 20th centuries. Until recently, the

putting together of useful gene packages by breeders – to give us our modern highly selected crop varieties, and commercial breeds of cattle, sheep and poultry – was carried out almost entirely by moving genes within species. Transfer of genes between species by classical genetic means, although it can sometimes be carried out in plants, is difficult – the hybrids generally have low fertility – and can only be done between closely related species.

With the great advances in DNA biochemistry in the last few decades of the 20th century, all this changed. It became possible to isolate the stretch of DNA corresponding to a particular gene from one organism and by various means transfer it to the genome of another species which might be quite unrelated to the first. In plants, the means of transference is sometimes the bacterium *Agrobacterium tumefaciens* which, being a plant pathogen, has the ability to invade plant cells and incorporate some of its own DNA into the plant chromosomes. The gene to be transferred can first be inserted into the bacterium and when that infects the plant cells, usually in tissue culture, the added gene becomes, in a few of the cells, incorporated into a plant chromosome. A cruder but surprisingly effective method is to coat tiny tungsten or gold particles with the DNA to be transferred, and shoot them at high velocity into plant cells in tissue culture. In a very small proportion of the cells so treated the DNA becomes incorporated into one of the chromosomes. The genetically transformed cell, regardless of which technique was used to insert the foreign DNA, can be selected out and induced to regenerate plants which can then be grown in soil and multiplied in the usual way. There are also techniques – e.g. direct microinjection into a fertilized egg, or using a virus to carry the DNA – for transferring genes into animal genomes, and there are yet other techniques for inserting foreign genes into the genomes of micro-organisms, particularly bacteria and fungi. In short, the species barrier which, since life began, has prevented transfer of genes between unrelated kinds of organism, has now been breached. Just about any gene, provided we know how to identify its DNA and isolate it, can be transferred to any plant, animal or microbial species.

In the case of agriculture one of the most significant contributions so far has been the creation of crop varieties resistant to particular herbicides so that, with the appropriate sprays, weed competition can be eliminated without harming the crop. Another major development has been the insertion into the genome of certain crop plants, particularly cotton, of the gene for a protein from *Bacillus thuringiensis*, which is toxic to certain insect pests such as cotton bollworm. In the pharmaceutical field, genetically transformed microbes are used to manufacture insulin and other medically important products.

Genetic engineering is clearly a very powerful technology. As with all powerful technologies – nuclear energy, synthetic organic chemistry – there are dangers as well as benefits. Such technologies can always be misused, usually through carelessness but sometimes with malevolent intent. It is therefore right and proper that people should be concerned about the possible unforeseen consequences of genetic engineering. The public expression of such concern, we may hope, will keep genetic engineers on their toes and make them continually examine, and seek to anticipate, the environmental and other consequences of the release of new organisms created in this way.

That being said, the fact is that genetic engineering has not yet given us a Chernobyl or a Bhopal. And although we must be ever-vigilant, there is no justification for the automatic, unthinking blanket rejection of all genetic engineering so often expressed by spokespersons in the 'Green' movement. While genetic engineering can have undesirable consequences, especially when taken up by unbridled capitalism devoid of a social conscience, it is much too valuable a tool, with tremendous potential for doing social and environmental good, to abandon. A recent example is the creation of a new form of rice which makes Vitamin A precursors in the grain: this 'golden rice' may one day significantly reduce vitamin A-deficiency blindness in the third world.[13] Insertion of genes for new proteins can change the amino acid composition of plant foods such as legumes, making them even better for human nutrition. Such improved

plant protein sources could also potentially substitute for the present wasteful use of fishmeal in aquaculture feedstuffs.

An example from viticulture highlights the potential environmental benefits from genetic engineering. At present, all vineyards have to be sprayed with fungicides many times each growing season to protect vines against downy mildew and powdery mildew, two very serious diseases to which the European vine, *Vitis vinifera,* is susceptible. Certain *Vitis* species from North America (where these diseases originated), however, are naturally resistant to the diseases. If the relevant genes in the American species could be identified and transferred to *V. vinifera* by genetic engineering, then grape varieties essentially identical to those we presently use, but resistant to those two devastating diseases, could be created and the amount of spray chemicals discharged to the environment during viticultural operations would be greatly reduced.

Plants already provide us with all kinds of useful chemicals: not only pharmaceuticals, but also bulk feedstock chemicals for industry, such as the 22-carbon erucic acid present in the older varieties of rapeseed, which is used as a precursor in the synthesis of a certain type of nylon. With the help of genetic engineering there are unlimited possibilities for persuading plants to make chemicals that will benefit human society. We could also greatly improve plants as sources of energy. Plants already use solar energy to accumulate chemical energy in the form of carbohydrate. There is no reason why varieties should not be developed in which a large part of this energy storage is switched to some forms of hydrocarbon, thus providing us not only with useful chemical feedstocks but also with liquid hydrocarbon fuel for transport purposes – a fuel derived from a solar, not a fossil, source.

I have given just a few randomly selected examples. But the beneficial possibilities of genetic engineering are endless and, while we must remain constantly vigilant about possible unforeseen side-effects, it would be an enormous pity if the human race, through timidity, was to abandon the use of this powerful tool.

POSTSCRIPT

I have argued in this chapter that we need science both to understand, and to solve, our environmental problems. Indeed it is our only hope. But science alone is not enough. Governments must come to grips with the scientific issues, and develop and implement the appropriate policies. This can be done. The CFCs and the ozone layer provide a good example. On the basis of evidence presented to them governments around the world accepted that the ozone layer was indeed under threat, and that chlorofluorocarbons and halons were the agents responsible; in 1987 they signed the Montreal Protocol which eventually led to the cessation of production of these compounds. The fact that there was cooperation from the scientifically aware chemical industries in the countries of production facilitated the establishment of this crucial international agreement. On other environmental issues, however, especially where powerful industries with a commercial axe to grind lobby effectively in the corridors of power, governments can drag their feet. This is exactly what is happening in relation to measures to curb greenhouse gas emissions, where two of the main culprits seeking to delay implementation of the Kyoto Agreement are the US and Australian governments.

All the science in the world will not save us if the political will to act is lacking.

Chapter 7

Gaia

Gaea … Greek personification of the Earth as a goddess … Gaea may have been originally a mother goddess worshipped in Greece before the Hellenes introduced the cult of Zeus. Less widely worshipped in historical times, Gaea was described as the giver of dreams and the nourisher of plants and young children

Encyclopaedia Britannica, 15TH EDN, VOL 5 (1993)

WITHIN THE GENERAL AREA of environmental concern there is one particular idiosyncratic school of thought, associated with British scientist James Lovelock, which draws its inspiration directly from science and whose implications are so far-reaching as to warrant treatment on its own. The central thesis might be expressed as follows.

We know that individual organisms behave in ways that favour their own survival. We know that groups of organisms – flocks of birds, schools of fish, colonies of insects – can also exhibit behaviour which favours survival of the group as a whole. Let us now make a giant conceptual leap and ask the question – what about the totality of living organisms on this Earth, the biota: can this be said to behave in such a way as to ensure its own survival through geological time? According to Lovelock, the answer is, Yes.

Origin of the Gaia Hypothesis

Back in the 1960s the US space agency NASA was preparing for the Viking mission to Mars. NASA enlisted the help of Lovelock, a well-known atmospheric scientist, to devise tests for the presence of life on that planet. He suggested that a good way to detect life on any planet would be a careful analysis of the composition of its atmosphere. A lifeless planet would have an atmosphere determined by physics and chemistry alone, with a composition close to chemical equilibrium. Where life-forms are present, on the other hand, they would be expected to betray their presence by their effects on atmospheric composition, resulting from their utilization of elements within, and discharge of wastes to, the atmosphere.

These considerations led Lovelock to realize how very far from chemical equilibrium the atmosphere of Earth is compared to those of Mars and Venus, which turned out to consist mainly of carbon dioxide and are indeed at chemical equilibrium. On our planet, life does not leave mere traces in the atmosphere but is in fact a major determinant of its composition. For example, the Earth's atmosphere contains a high concentration of oxygen and a low concentration of carbon dioxide because, during photosynthesis, plants remove CO_2 and liberate oxygen, and they have been doing so on Earth for the past $3\frac{1}{2}$ billion years or so. The universal presence of the highly oxidizable gas, methane, in the Earth's atmosphere despite the presence of 21% oxygen, is another sign of the presence of life on this planet.

The story as I have outlined it so far is uncontroversial, and is broadly accepted. But Lovelock went further. He was struck by the fact that not only is the composition of the Earth's atmosphere far from chemical equilibrium, but geological evidence indicated that it was stable over periods of time which were very long compared to the residence time of the gases present. This led him to conclude that there must be a control system responsible for maintaining this approximate constancy.

Another aspect of the Earth's history which seemed to call for some kind of control system was its surface temperature. Astrophysical considerations

indicate that 3–4 billion years ago, when life is thought to have begun, the Sun emitted about 25% less energy than it emits today and that its luminosity has been steadily increasing to the present level during that time. Carbon dioxide is continuously being discharged into the Earth's atmosphere from volcanoes; it has been proposed that the planet was kept from freezing over in that far distant era by a super greenhouse effect caused by the high levels of CO_2 believed to be present in the atmosphere then. The question arises – given the major increase in solar luminosity which has taken place, why did the Earth not become overheated (like Venus) and thus unsuitable for life? To put it another way, how has it come about that over an enormously long period the Earth's surface temperature has remained within the narrow range that permits the existence of life?

Major atmospheric and climatic control systems on this planet must be either physicochemical or biological, or a combination of the two. To assist us in our consideration of the biological contribution, it is convenient to make use of the concept of the *biosphere*. We may define this as that layer of air, earth and water surrounding this planet, within which living organisms have their being. The biosphere is the whole thing – the thin spherical shell of inhabitable space, together with the organisms within it. In his quest for the control system responsible for the relative constancy of atmospheric composition and surface temperature, Lovelock eventually concluded that it was primarily biological in nature; that it was the biosphere itself. In collaboration with evolutionary biologist Lynn Margulis, he arrived at an explicit formulation of the hypothesis:

> *Life, or the biosphere, regulates or maintains the climate and the atmospheric composition at an optimum for itself.*

This hypothesis was arrived at, as we have seen, on the basis of Lovelock's response to the enormous differences between the climate and atmosphere of Earth and those of the lifeless planets, together with the apparent stability of environmental conditions on our planet over very long periods. A hypothesis of this type can legitimately be put forward simply on the

basis of the information we have about the terrestrial-type planets of the solar system, even if no-one can yet think of mechanisms to explain how such a regulatory system works. It is important to realize that such a hypothesis remains scientifically legitimate, and could even be true, regardless of whether particular mechanisms put forward to explain its inner workings turn out to be correct. The fact that the human body possesses control mechanisms which control its temperature and internal chemical environment to a very high degree of precision, was realized long before it became possible to ascertain how those mechanisms actually work.

Possible Gaian Control Mechanisms

Lovelock and his co-workers have proposed some possible mechanisms as to how planetary control mechanisms of a biological nature might operate. When photosynthetic life forms first emerged ~3.5 billion years ago in the sea they used light energy to convert carbon dioxide into biomass. While most of this biomass would be broken down to liberate CO_2 again, together with methane in the anaerobic conditions prevailing in that far-off era, nevertheless every year a small proportion of the annual production of organic matter would sink to the bottom of the ocean and be removed from circulation. This 'biological pump' is a well-known process and still operates today. In addition to removing carbon in dead plant biomass, some of the photosynthesis carried out by these primitive algae would be accompanied by precipitation of calcium carbonate out of solution. This too would sink to the bottom of the ocean, taking carbon with it. Since the carbon dioxide for oceanic photosynthesis comes ultimately from the atmosphere, these two processes provided a mechanism by which atmospheric CO_2 was steadily removed and its concentration progressively reduced. As the atmospheric CO_2 concentration diminished – eventually to the comparatively low level that has prevailed for the last few million years – the greenhouse effect diminished with it. In this way, it is proposed,

the increased heating effect of the ever more luminous Sun was counteracted and the Earth's surface temperature maintained within a range tolerable for life.

To be efficient and to avoid overshoot a control mechanism of this type needs to be able to respond continuously to the level of the global parameter being controlled. Otherwise it might become just a one-way mechanism for removing virtually all atmospheric CO_2, leading to excessive cooling and a fall in temperature which might be catastrophic for the biota. A possible negative feedback mechanism would involve transfer of mineral nutrients from the land to the sea. Growth of phytoplankton (the floating microscopic plant cells) in the ocean is mostly limited by the availability of the essential nutrients, nitrate and phosphate. When global temperatures are high there is more evaporation from the sea and consequently more rain on the land, leading to increased leaching of nutrients from soils into the rivers and then into the sea, thus promoting phytoplankton growth. As the Earth cools, due to falling atmospheric CO_2 levels, rainfall diminishes, transfer of nutrients from land to sea is curtailed and phytoplankton production diminishes, so the biological pump transferring carbon from the atmosphere to the deep sea slows down. The continuing venting of CO_2 to the atmosphere from volcanoes, combined with the decreased rate of removal, would then ensure that the atmospheric CO_2 concentration does not fall to dangerously low levels.

It has also been proposed that marine phytoplankton can lower global temperature through an effect on cloud formation. The major source of cloud-condensation nuclei (CCN) over the oceans is an aerosol of sulphate particles, most of which are formed by oxidation of dimethyl sulphide, a compound derived from phytoplankton cells in the sea. Robert Charlson, James Lovelock and co-workers suggest[2] that as phytoplankton populations increase, more CCN are generated, clouds increase and more solar radiation is reflected back to space, so the Earth cools. They estimate that an approximate doubling of CCN would be needed to counteract the warming due to a doubling of atmospheric CO_2.

A possible additional negative feedback mechanism is provided by the fact that phytoplankton growth is limited not only by the level of mineral nutrients but also by the actual seawater concentration of carbon dioxide, the carbon source for the photosynthetic process. The concentration of CO_2 in the surface layer of the ocean, where most of the photosynthesis goes on, is closely linked to that in the atmosphere. Thus as the atmospheric CO_2 level falls the seawater concentration also falls, putting a further brake on phytoplankton photosynthesis.

In photosynthesis, light energy is harnessed to take the hydrogen from the water molecule, H_2O, and use it to reduce CO_2 to the level of carbohydrate and other cellular constituents. The oxygen component of the water molecule is simultaneously liberated as molecular oxygen, O_2. In nature, most plant biomass produced in this way is broken down by micro-organisms when the plant dies and reconverted to CO_2 and H_2O, an oxidative process which consumes an amount of oxygen equivalent to that liberated in the plant's original photosynthetic growth. However, the loss to deep ocean sediments of a small part of the organic matter formed by photosynthesis each year means that some of the photosynthetically liberated oxygen remains unconsumed. It is now believed that this is the mechanism by which the level of oxygen in the atmosphere has risen from zero when life first began to appear (and for perhaps more than a billion years after that) to its present high level of 21%.

This accumulation of oxygen, since it made possible the evolution of active creatures such as ourselves, may seem like a good thing but in fact it presents a very serious problem since the flammability of organic matter increases sharply with oxygen concentration. At 25% atmospheric O_2 (a mere 4% increase) even damp tropical rainforest would burn very readily, and it is difficult to see how any terrestrial vegetation would survive.[3] In fact, oxygen concentration is prevented from rising to these dangerous levels by the operation of another part of the biosphere, namely the anaerobic bacteria of the marshes, wetlands, estuaries and ocean bed. Their non-oxidative breakdown of plant matter liberates in excess of 1000 mega-

tons of methane every year. This finds its way into the atmosphere where it undergoes oxidation back to CO_2 and water, in the process consuming correspondingly vast amounts of oxygen. Lovelock proposes that there are control mechanisms, involving the methane production system and perhaps also terrestrial forests, responsible for the surprising constancy of atmospheric oxygen at ~21% for at least the last 200 million years.

Another possible example of the biosphere influencing a major global chemical process is the relationship of oceanic phytoplankton to the cycling of the element, sulphur. This is plentiful in the sea but in short supply on land. Marine phytoplankton contain the sulphur compound dimethylsulphoniopropionate (DMSP). When the phytoplankton cells are consumed by zooplankton,[1] the DMSP is broken down to the volatile sulphur compound dimethyl sulphide, which escapes to the atmosphere where it is oxidized to sulphate then washed down in rain onto the land where it can be used by terrestrial plants. Marine algae excrete volatile organic iodide compounds, such as methyl iodide, which escape into the atmosphere, where they are converted into iodine oxides and iodine oxyacids[8] which are also carried by rain down onto the land. Calculations by Peter Liss show that these are the main mechanisms by which sulphur and iodine are transferred from the sea to the terrestrial biosphere.

If James Lovelock is right then the biosphere, considered as a totality, is regulating its internal environment in such a way that life within that internal environment, and therefore the biosphere as a whole, remains active and healthy. This is exactly what higher organisms such as ourselves do. We have various feedback mechanisms which ensure that our internal environment remains in a state such that we continue to function. We have mechanisms which control our body temperature, the electrolyte balance of our blood and so on.

The Lovelock hypothesis, proposing that the biosphere has feedback mechanisms regulating its internal environment so that it can cope with an increase in solar luminosity or whatever, is thus analogous to regarding the biosphere, or indeed the planet Earth, as a living organism. This idea –

that the Earth is in many ways similar to a living organism – is quite an old one. It was expressed particularly clearly by Russian scientist Vernadsky in the early decades of the 20th century, and in the 18th century by James Hutton, the founding father of geology. Coincidentally, James Lovelock lived in the same English village as William Golding, the Nobel Prize-winning novelist. Golding suggested to Lovelock that anything which is alive deserves a name; he suggested 'Gaia', the Greek name for the Earth goddess. Thus the hypothesis we have been discussing came to be called the *Gaia* hypothesis.

More prosaically, the Gaia hypothesis might be regarded as an example, on a gigantic scale, of the well-known *Le Chatelier principle* of physical chemistry. This principle, formulated by Henry Le Chatelier in 1888, states[7] that:

> *Any change in one of the variables that determine the state of a system in equilibrium causes a shift in the position of equilibrium in a direction that tends to counteract the change in the variable under consideration.*

Many examples are known in simple systems. For example, if a particular chemical reaction in the gaseous state is accompanied by a decrease in volume, then if the pressure applied to the gaseous mixture at chemical equilibrium is increased the equilibrium is shifted in the forward direction, i.e. more of the gases undergo reaction, thus reducing pressure and so tending to nullify the increase in applied pressure.

CRITICISMS OF THE GAIA HYPOTHESIS

The Gaia hypothesis as stated by James Lovelock and his colleagues such as Lynn Margulis and marine scientist Andrew Watson is, whether right or wrong, of an entirely legitimate scientific type. It has, however, met a mixed reception within science. Some of the criticisms have been of a perfectly proper scientific nature, but others seem to reflect more philosophical or

even ideological concerns. For any theory to be accepted as scientific, it must in principle be possible to refute it. That is, it must be possible to devise some test which can be applied to the theory, such that if the theory fails the test then it would be shown to be wrong. It is generally accepted that by some means the Earth's surface environment has been maintained in a state suitable for life for billions of years. The central proposition of the Gaia theory is that it is primarily the activities of the living organisms themselves which have achieved this. If it could be shown that purely physicochemical processes would have achieved this on their own, without significant contribution by the biota, then the Gaia theory would be refuted. Geochemist James Walker and his colleagues have sought to explain the progressive diminution of atmospheric CO_2 over the last few billion years, which saved this planet's life-forms from a terminal greenhouse effect, by a purely inorganic mechanism. It is already known that weathering of silicate minerals on the land by rainwater containing carbonic acid from dissolved CO_2 brings about a net transfer of carbon from the atmosphere to the deep sea (see Box, p. 211). As the Earth warmed up due to the increasing luminosity of the Sun, evaporation from the ocean increased. This led to increased rainfall and consequently more weathering of silicate minerals and transfer of atmospheric CO_2 to the deep sea. Earth's surface temperature would then have fallen, due to the decreased greenhouse effect.

This is a good, clear simple theory, and if it did the job it would be Goodbye Gaia. It should be emphasized that the fact that the Gaia hypothesis can be put to such a test is a point in its favour, indeed essential if we are to regard it as a proper scientific theory. The indications so far are that Gaia is likely to survive this test. Lovelock accepts that transfer of carbon by this weathering process does occur, and agrees that such weathering has played a major role, but he believes that in the form envisaged by Walker it is far too slow to have achieved the reduction in atmospheric CO_2 that has actually occurred. He argues that sufficiently high rates of weathering are achieved only in the presence of living organisms, and it has now been

demonstrated[5,12] that the rate of weathering of silicates is accelerated manyfold by living organisms. The mechanism seems to be that the transfer of large amounts of organic matter to the soil by plants, and its breakdown by micro-organisms, vastly increases the concentration of CO_2 within the soil's pores and in this way greatly increases the rate of reaction with silicate minerals (see Box, p. 211). In effect, the Lovelock argument is that certainly weathering of silicates is one important mechanism for pulling down atmospheric CO_2 levels and therefore temperature, but it needs the biota in order to work efficiently and is thus part of the Gaia system.

The proposal that marine phytoplankton can affect global climate by increasing sulphate-containing cloud-condensation nuclei, and thus promoting cloud formation, has also been criticised. Stephen Schwartz[11] argues that if this is true then anthropogenic release of sulphur dioxide into the atmosphere over the last century and more in the northern hemisphere should have had the same effect. He believes that 100-year temperature records and current cloud albedo (reflectivity) data do not support this hypothesis. James Lovelock regards this particular proposition of Gaia theory as being still under test.[6]

A more general criticism levelled at the Gaia hypothesis is that it is teleological, *teleology* being a view or doctrine which asserts that developments in the world are due to (caused by) the purpose or design which they serve. Lovelock stoutly denies this:

> *Neither Lynn Margulis nor I have ever proposed a teleological hypothesis. Nowhere in our writing do we express the idea that planetary self-regulation is purposeful, or involves foresight or planning by the biota.*

The apparent cooperation by the different living elements of the biosphere to achieve a stable and liveable environment for all, implied in Gaia, seems to imply some form of altruism. The theory was criticized by Richard Dawkins on the ground that Darwinian natural selection provides no mechanism by which altruism on a global scale could evolve. Lovelock and Watson responded by devising an imaginary model planet, 'Daisyworld',

populated by two competing species of daisies, a dark-coloured one which could absorb solar energy and cause its surroundings to heat up, and a light-coloured one which would reflect solar energy and thus tend to cool its surroundings. Numerical modelling showed that changes in the relative proportions of the two species, governed by natural selection, made it possible for this simple imaginary system to achieve effective temperature regulation and keep its climate constant over a wide range of increasing solar luminosity.

That a self-controlling global ecosystem could arise naturally is thus not necessarily incompatible with Darwinian natural selection. Nevertheless, it must be said that the reason Daisyworld works is that the properties of the competing species were carefully chosen so that it would work. It is not obvious that the collective properties of the myriad prokaryote and eukaryote species that have come and gone on this real planet are such that, in the kind of physicochemical conditions that actually exist here, they would achieve the control and stabilization of the environment that the Gaia hypothesis gives them credit for. However, the observational case for the operation of such control, while not overwhelming, is highly persuasive. It may be that we will eventually just have to accept, in the same spirit in which we contemplate the fact that the physical constants of the universe seem to be precisely as we would wish them to be, that once again we have been lucky: that the emergence of such control is, for reasons we do not yet understand, in some way built into any global carbon-based ecosystem that could ever arise on a planet such as ours. In the meantime, the search for biogeochemical mechanisms by which the proposed Gaian control could be achieved, must continue.

Implications of Gaia for environmental concern

What are the implications of Gaia for environmental concern? They are rather mixed, and when properly understood they will not please everyone

in the environmental movement. The Gaia concept has certainly been taken up with enthusiasm by environmentalists because it portrays this beautiful planet, about whose future welfare we are all so concerned, as a living being, indeed almost as a person. Human nature is such that we can relate to individual creatures, especially individual people, more readily than to complex concepts such as the totality of natural ecosystems. The very name 'Gaia', portraying our planet as a sort of Earth mother goddess, has excited all kinds of good resonances.

However, on reflection, some rather disturbing implications can be seen. The central proposition of the Gaia hypothesis is that the biosphere acts to maintain itself using an array of feedback mechanisms. But it is the well-being of the biosphere as a whole which is important, not necessarily any individual species, or even any particular ecosystem. Lovelock points out that certain specific ecosystems which are particularly active in controlling the physicochemical properties of the biosphere may be crucial to the welfare of Gaia. He believes likely candidates are estuaries, the continental shelf, wetlands, the humid tropical forests. But other ecosystems – even some of which we human beings might be very fond – may be largely irrelevant to the welfare of the biosphere as a whole, and by that standard their destruction would be of little significance.

Loking at wildlife conservation from a Gaian perspective, Gaia has seen, through evolutionary time, a continuous turnover of species. The disappearance of many wildlife species from the Earth may be of little significance for the continued survival of the biosphere as a whole, especially when you realize, as Lynn Margulis has pointed out, that many of the important species in the biospheric regulation process are unglamorous micro-organisms. And if there is one species that Gaia would certainly never miss, it is *Homo sapiens.*

Thus, from an environmental point of view the Gaian perspective reveals the importance of identifying and protecting those ecosystems which are crucial for the continued functioning of the biosphere. It does not, however, provide any broad justification for ecosystem and wildlife conser-

vation other than the purely pragmatic one that any given ecosystem may just possibly turn out to play an important role in biospheric regulation. But that will very often be an unconvincing line of argument. So, let the environmental movement by all means learn something from the Gaia hypothesis, but it should not expect the hypothesis to provide all the intellectual ammunition needed for the causes dearest to its heart.

REMOVAL OF ATMOSPHERIC CO_2 BY WEATHERING OF SILICATE MINERALS

Calcium silicate ($CaSiO_3$) is a very common component of rocks, and therefore of the soil derived from those rocks. Rainwater draining through soil contains elevated levels of CO_2 from microbial breakdown of plant-derived organic matter in the soil. The dissolved CO_2 reacts with water to give carbonic acid (H_2CO_3). This weak acid dissolves the calcium silicate in the soil particles by reacting with it to give calcium bicarbonate ($Ca[HCO_3]_2$) and meta-silicic acid (H_2SiO_3):

$$2H_2O + 2CO_2 \rightarrow 2H_2CO_3$$

$$CaSiO_3 + 2H_2CO_3 \rightarrow Ca(HCO_3)_2 + H_2SiO_3$$

The overall reaction is thus:

$$CaSiO_3 + 2H_2O + 2CO_2 \rightarrow Ca(HCO_3)_2 + H_2SiO_3$$

The freely soluble calcium bicarbonate leaches out of the soil in the drainage water, and is conveyed by rivers to the sea. In the sea, plankton use calcium bicarbonate as a source of calcium carbonate ($CaCO_3$) for shells and other structures. When the organisms die this calcium carbonate falls to the sea floor, thus bringing about the final stage in the transference of atmospheric CO_2 to the bottom of the sea.

Chapter 8

'-isms'

We are born believing. A man bears beliefs as a tree bears apples.
Ralph Waldo Emerson, *Conduct of Life: Worship*

Every man, wherever he goes, is accompanies by a cloud of comforting convictions, which move with him like flies on a summer day
Bertrand Russell, *Sceptical Essays*

All human beings have have systems of belief. It is the possible implications of science for the validity of those beliefs which give meaning to our existence that grate most painfully on human sensibility. It is true that some systems of belief are not compatible with the scientific world view. It is also, however, true that the claims of certain other belief systems to derive a special legitimacy from science are equally invalid. We shall examine some of these problems here. Our final question is – what does science have to say about our most cherished beliefs?

Theism and atheism

Theism is a philosophical doctrine which holds that the physical universe does not contain within itself the explanation of its own existence: that it was created, and is constantly sustained, by a non-material entity existing

outside time and space. This entity or being is commonly referred to as 'God'.

Atheism is the opposite of theism. In its most common form – materialist atheism – it asserts that nothing exists but the observable physical universe. It follows, although this is not always explicitly stated, that the universe must contain within itself the explanation of its own existence.

Science cannot adjudicate between theism and atheism. The surprisingly common assumption – so common and so uncritically accepted that it must be one of Richard Dawkins's 'memes'[11] – that science somehow provides special support for the atheist position, has no basis. Science takes the existence of the physical universe as a given, and proceeds from there. Although science can provide a plausible, but still very hypothetical, account of how the universe could have arrived at its present state, from what we believe to have been its original very different state ~13.7 billion years ago, it cannot tell us why there should be a universe in the first place, and why it should contain these particular physical laws and not some others, or indeed any laws at all. Such questions about existence, origins, meaning and purpose are *metaphysical* not physical. They have traditionally, and rightly, been the concern of philosophy and religion. This is not to say that scientists are not as entitled as any other human beings to hold opinions on these matters. But in their capacity as *scientists* the most they should do is to urge that any metaphysical theories concerning the nature of existence should at least be consistent with the picture of physical and biological reality that scientific investigation has so painstakingly built up over the centuries.

Although science cannot resolve the theism/atheism debate, it can provide information about the natural world which might dispose people to find one or the other side more plausible. How, for example, does the world view presented by physics line up with the metaphysics? Physics tells us that underneath the enormous complexity confronting us in the real world, there lies fundamental order and simplicity. There are a comparatively small number of physical laws governing the behaviour of the

universe and these can be expressed in terms of simple, elegant mathematical equations. Many scientists have found it very surprising both that the universe is at heart so simple, and that evolution has given us minds with the capacity to understand it. The eminent physicist Eugene Wigner (Nobel Prize, 1963) speaks of the 'unreasonable effectiveness of mathematics' in revealing the structure of the physical world.

A physical description of the world consists not only of mathematical laws, but also of a set of physical constants, numerical values that certain fundamental physical quantities – the speed of light, the mass and charge of the electron and the proton, the fine-structure constant, the cosmological constant etc. – actually possess. It turns out that all kinds of crucially important physical processes – important for the formation of galaxies, stars, the chemical elements, complex molecules, living organisms – are dependent on the values of these physical constants. This is not surprising in itself. What is surprising is the extraordinary degree of precision with which the values of these constants must be specified if the physical universe is to be one in which life could evolve. An encyclopaedic treatment of this topic is given in *The Anthropic Cosmological Principle* by physicists John Barrow and Frank Tipler. Just a few examples will suffice us. As well as the proton (hydrogen nucleus) and the deuteron (proton + neutron, helium nucleus), another nucleon (the diproton) consisting of two protons might conceivably exist. The actual value of the nuclear strong coupling constant is such that the diproton only just fails to exist. To quote Barrow and Tipler:[2]

> *The existence of deuterium and the non-existence of the diproton therefore hinge precariously on the precise strength of the nuclear force. If the strong interaction were a little stronger the diproton would be a stable bound state with catastrophic consequences – all the hydrogen in the Universe would have been burnt to ^{2}He during the early stages of the Big Bang and no hydrogen compounds or long-lived stable stars would exist today. If the di-proton existed we would not!*

Furthermore, if the strong interaction was slightly weaker only hydrogen would exist; heavier elements would not form. As eminent theoretical physicist Freeman Dyson puts it:[14]

> *If, as seems likely, the evolution of life requires a star like the sun, supplying energy at a constant rate for billions of years, then the strength of nuclear forces had to lie within a rather narrow range to make life possible.*

The values of two other fundamental physical forces – gravity and electromagnetism – also show signs of very fine tuning. A balance between these two forces is required for stars to be stable objects. According to astrophysicist Brandon Carter, as quoted by Paul Davies,[10] changes in the strength of either force by only one part in 10^{40} would spell catastrophe for stars like our Sun.

All life, conceivable as well as actual, is based on compounds of carbon (I confidently disregard all the chemically and physically implausible speculative fantasies about other possible bases for life). Carbon is formed by nuclear reactions within stars. This is only made possible by another fortunate 'coincidence'. We saw earlier (Chapter 1) that carbon is synthesized by the nuclear reactions:

$$^{4}He + {}^{4}He \rightarrow {}^{8}Be$$

$$^{8}Be + {}^{4}He \rightarrow {}^{12}C + 2\gamma$$

It was predicted by astrophysicist Fred Hoyle that, for this reaction to occur at any significant rate, it is necessary for the carbon nucleus to have a resonance level of about 7.7 MeV, just above the sum of the energies of ^{8}Be and ^{4}He. Experiments shortly thereafter proved this prediction to be correct, an appropriate level being found at 7.656±0.008 MeV. By another fortunate coincidence, the carbon once formed only just escapes being 'burned' to oxygen by the reaction:

$$^{12}C + {}^{4}He \rightarrow {}^{16}O$$

because the nearest resonance level of the oxygen nucleus is just below, rather than just above, the sum of the energies of ^{12}C and ^{4}He.

Life not only requires carbon compounds, but life processes only go on in a medium consisting predominantly of water. Water is the most peculiar liquid in the universe, with a host of anomalous properties, without which life could not exist. This was realized as long ago as the 19th century, and was explored in detail by Harvard chemist Lawrence Henderson in the early years of the 20th century. He pointed out, *inter alia*, that one well-known such property – the fact that water expands, thus becoming less dense, on freezing – is essential for life. If ice were denser than water all surface water bodies would freeze solid from the bottom up, killing the aquatic biota. As it is, ice forms a protective insulating layer on the surface, below which the water, although cool, remains liquid and suitable for life.

The realization that the universe has these surprising properties, that it is constructed with the utmost precision in such a way as to make possible the emergence of life, and eventually human beings, has led many scientists, particularly in the physical sciences, to the strong suspicion that the universe is the product of design, not accident. The well-known astronomer Fred Hoyle was so impressed with the series of extraordinary coincidences required in the laws and constants of nuclear physics to make possible synthesis of the element carbon (upon which all organic life is based) in the stellar core, that he was led to wonder if the laws of nuclear physics have been deliberately designed with regard to the consequences they produce inside stars. He eventually arrived at the conclusion that the universe is, as he phrased it, ‘a put-up job’. The view of an earlier eminent astronomer, Sir James Jeans, was that the universe appears to have been designed by a pure mathematician. Freeman Dyson concluded ‘from the existence of these accidents of physics and astronomy that the universe is an unexpectedly hospitable place for living creatures to make their home in’.[14] The physicist Paul Davies, who has written extensively on this topic, concludes, in relation to the laws governing the physical origin of the universe, that they seem to be the product of exceedingly ingenious design. He goes on to say that:

If physics is the product of design, the universe must have a purpose, and the evidence of modern physics suggests strongly to me that the purpose includes us.

It used to be the case that physicists confidently expected that one day we would arrive at a grand 'theory of everything', which would explain not only all the physical laws but also why the variety of physical constants have the particular values they do. Confidence in the existence of such a theory, while it has not disappeared, is no longer universal. It seems that 'a growing number of cosmologists and string theorists suspect the form of our Universe is little more than a coincidence' and 'are increasingly turning to dumb luck as an explanation'.[4]

Rather than – God forbid! – propose God as an explanation, one reaction has been to propose that perhaps there are an enormous number of universes (anything from 10^{500} up to infinity), within which the physical constants and laws vary all over the place, and we just happen to find ourselves within one where the laws and constants are just right to permit the existence of intelligent life.[23,28] This is sometimes referred to as the multiverse hypothesis. The problem is that the hypothesis is inherently unfalsifiable: it cannot be put to experimental test. We have no evidence, and it is not easy to imagine how we could ever have evidence, for the existence of even one other universe, let alone 10^{500}. As for an infinity of universes, this is meaningless: infinities are mathematicians' playthings, not legitimate physical concepts.[16] Thus, the 'multiverse' is interesting metaphysics, but not part of science. We should note in passing that the multiverse hypothesis in no way precludes theism. A multiverse could be created just as easily as a single universe.

The search for a theory of everything is the ultimate expression of the belief that the universe is both fundamentally law-like and comprehensible by human beings. While most scientists today implicitly believe this without ever having given it much thought, it would by no means have been obviously true to early men and women, subjected as they were to random, unpredictable environmental catastrophes – flood, drought, forest fire,

pestilence etc – throughout their lives. There were promising beginnings of science and mathematics in a number of early cultures – Greek, Indian, Egyptian, Chinese, Mayan – but it was in Europe that science finally took off. Since we now know that people of all ethnicities are perfectly capable of doing science, the explanation has to be something other than a special virtue in Europeans. The brand of monotheism that universally prevailed in Europe, that proclaimed that the universe had been created by a single just and rational being, may have been crucial. With such a belief it was plausible to suppose that the universe should fundamentally make sense, and to seek to discover the God-ordained laws according to which it operated. To quote Isaac Newton:

> *The hypothesis on which I have based all my work in Natural Philosophy is that every natural law is God's Law, and therefore must express his Glory and Divine Purpose.*

What is perhaps surprising is that the automatic assumption that the universe must be fundamentally law-like is just as common among scientists who have lost the monotheism of their forebears as it is among believers. It is, of course, arguable that the operation of the universe in accordance with a certain finite number of laws, all expressible in mathematical form, can now be regarded just as an empirical fact, which has been revealed by the work of physicists and chemists over the last few centuries. That our laws of physics are either the 'right' ones, or at least very good working approximations of the right ones, is shown by the fact that they work: what they predict is what turns out to be the case. The fact that physical reality is to a large extent (but not yet wholly) explicable in this way is certainly our good fortune, but why it should be so, in particular why we find the 'unreasonable effectiveness of mathematics' in the real world, remains a mystery.

We have noted that an all-embracing physical theory has not yet been achieved, and the outlook is not promising. Before leaving this subject, I suggest that it would be a pity if physicists gave up the search for a theory

of everything. Their chances of success may not be good, but in the continuing attempt they are bound to make lots of interesting discoveries along the way. To travel hopefully is better than to arrive, or not to travel at all.

A view which still survives among some biological scientists is that any argument for theism from design is invalidated by the Darwinian theory of evolution, which proposes an explanation for the precise adaptation of living organisms to their environment entirely in terms of natural selection of variants generated by mutation, i.e. without recourse to the hypothesis that species are the product of design. To the extent that it is correct, this view of course only applies to living organisms. It is entirely irrelevant to the question of the evidence for design in the physical universe.

It should be pointed out that even this argument against design in the biological world is only valid if the simple neo-Darwinian mechanism for evolution is in fact correct. If the alternative view, discussed in Chapter 2 (that there are a range of possible natural structures, with a tendency to arise spontaneously, and that Darwinian natural selection simply operates on and selects among these) is correct, the spectre of intelligent design may come again to haunt the halls of evolutionary biology. The implication would be that not only are the laws of physics and chemistry such as to make possible the existence of complex, carbon-based biological polymers, but that these polymers have a built-in, arguably designed-in, propensity to come together in highly specific ways to form certain kinds of structures and not others.

A particularly serious stumbling-block for materialist atheism is the existence of conscious mind. No matter at what scientific level we look at it, whether the integrated operation of the whole brain/nervous system, the electrochemical physiology of neurons, the detailed biochemistry of synaptic processes, the quantum mechanical influences on synaptic vesicle discharge or tubulin conformation change, at no point do we see any reason why there should be such a thing as conscious thought. It cannot be found in the fundamental equations of theoretical physics by which we assume all brain processes are ultimately controlled. Nothing in the func-

tioning of the enormously complicated human brain, as revealed to us by the whole armoury of physicochemical techniques which have been applied to it, would lead us to suppose that, associated with the structures and processes that we observe, there should arise anything so alien to the physical world, so utterly different, as consciousness. That there should simultaneously exist the physical world, and the world of conscious thought, is not merely *a* problem. It is *the* problem.

The existence of these two worlds does not, of course, prove that theism is right and atheism is wrong. Its significance is that materialist atheism derives much of its persuasive power by combining the undoubted success of science in explaining how the physical world works (although not why it or its laws should exist) with the tacit assumption that nothing but the physical world exists. Consciousness has no place on this world view, and its undoubted existence represents a serious problem for materialist atheism. For theism, on the other hand, the existence of both a physical world and the world of conscious thought presents no problem at all.

In summary, each of the two claimants, theism and atheism, for our belief is a bold metaphysical conjecture. Neither can be falsified by science. Equally, neither position gets any special deal from science. What can be said is that people who are open to the possibility of there being design in the universe can find much in the physical laws and constants of the universe to support that view.

Theism/atheism and environmentalism

Before leaving theism and atheism it is instructive to consider the implications of each of these opposed metaphysical positions for that other major area of human sensibility upon which science has a bearing, namely, environmental concern.

We saw in an earlier chapter that a fundamental principle espoused by many environmentalists, especially the adherents of deep ecology, is that

living species have inherent value independent of their usefulness for humanity. Those who accept this view commonly take it to imply that we human beings, since we now control so much of what happens on this planet, have a moral obligation towards other species; in particular, we should try to ensure that our activities do not drive any of them to extinction. Unfortunately the assertion that species have intrinsic value in their own right derives no more support from science than any other attribution of value. As environmental philosopher J. Baird Callicott puts it:[5]

> *At the level of organization with which we are concerned, the macroscopic world of terrestrial life and the value of its component species, the classical attitude that nature is value-neutral remains a virtually unchallenged dogma of the scientific world view. From this perspective, the attribution of intrinsic value to species, as to anything else under the sun, is doomed at the outset to failure.*

This does not mean that all is lost. We do not need science's permission to talk about value. What it does mean is that if we seek intellectual justification for attributing value to non-human species, or to anything else, we must look beyond science. How about to our two metaphysical systems, theism and atheism? Atheism does not help, since it says in effect that there is only the scientific world view and nothing else. This is unfortunate for the many members of the environmental movement whose philosophical position is, if not always clearly thought out, essentially of the atheist variety. Theism, on the other hand, offers a great deal of scope for attributing value to nonhuman species, and perhaps even to whole ecosystems. The palaeontologist Simon Conway Morris refers[8] to:

> *the recurrent inability of materialists to understand that the decision to protect the biosphere can only derive from an ethical imperative that is itself independent of the natural world.*

J. Baird Callicott has explored the question of whether, among the various ethical systems in the philosophical market-place, any metaphysical

foundations for the intrinsic value of other species can be found. He concluded that there are essentially four distinct moral metaphysics which might yield a positive answer to this question. He refers to these as J-theism (a strand of thinking within Judaeo-Christian theism), holistic rationalism, conativism and bio-empathy. Callicott states that 'only J-theism unequivocally provides for objective intrinsic value for existing nonhuman species'. However, he believes that an environmental ethic may also be found in bio-empathy, a position he derives from Hume and Darwin, which simply acknowledges:

> *that the non-utilitarian value of other forms of life is ultimately emotional, that it rests upon feeling, that species are valuable and we ought to save them because we have an affection for them.*

Callicott does not think there is much mileage in the other two metaphysical systems, and since we do not want to get into philosophical technicalities we shall not discuss them here.

Theism holds, as we saw earlier, that the universe does not contain within itself the explanation of its own existence and nature, but was created: the creator is the being we refer to as 'God'. Everything in the universe belongs to and is maintained in existence by God. This of course includes the vast proliferation of life-forms on this planet, which came into existence by the creative process of evolution in accordance with the laws of physics and chemistry, built into the universe by its creator. According to the Genesis story of Judaeo-Christian theism, after creating the world:

> *God saw everything that he had made, and behold it was very good.*

While we, like every other living creature, may use other life-forms, the species themselves are not – according to this view – ours to destroy: they are part of the unfolding beauty and complexity of God's creation, and it is wrong for us to diminish that beauty and complexity.

While this may seem to follow plausibly from the Judaeo-Christian theistic position, it must be said that Christian churches were rather slow to

recognize this. In recent years, however, there has been a general awakening in the Christian community to the moral significance of conservation of the natural world (e.g. *To Care for the Earth: A Call to a New Theology* by Fr Sean McDonagh)[19]. In his 1990 World Day of Peace Message, a document with a strong ecological theme, the late Pope John Paul II referred *inter alia* to the 'lack of due respect for nature', 'widespread destruction of the environment', 'indiscriminate application of advances in science and technology', 'delicate ecological balances' being 'upset by the uncontrolled destruction of animal and plant life', the need for 'an education in ecological responsibility', the 'deep restorative power' of 'our ... contact with nature'. He stated clearly that 'the ecological crisis is a moral issue'. In an address[22a] to farmers from around the world, assembled in Rome in 2000, John Paul II said:

> *If the world of the most refined technology is not reconciled with the simple language of nature in a healthy balance, human life will face ever greater risks, of which we are already seeing the first disturbing signs*

and

> *Work in such a way that you resist the temptations of a productivity and profit that are detrimental to the respect for nature. God entrusted the earth to human beings 'to till it and keep it' (cf Genesis 2:15). When this principle is forgotten and they become the tyrants rather than the custodians of nature, sooner or later the latter will rebel.*

The Pastoral Statement of the US Catholic bishops in 1991, on the theme 'Renewing the Earth', stated that desirable objectives were 'to safeguard endangered species' and 'to preserve remaining wilderness'. It explicitly acknowledged the important concept of stewardship with the words 'We shall be required to be genuine stewards of nature'. Towards the other end of the Christian spectrum, a group of senior evangelical leaders in the US have released a document[29] calling for action on global warming.

People who accept Judaeo-Christian theism as a fundamentally correct world view and who are of an environmentalist disposition, especially those who happen to be Catholics (eco-Catholics?), now have a clear mandate as well as a coherent philosophical base for the pursuit of environmentalist ends, including the metaphysically problematic matter of conservation of biodiversity. Adherents of the atheistic world view but of a similar environmentalist disposition have something of a problem if they wish to ground their devoutly held conservationist beliefs in something more substantial than warm feelings: for the time being they should perhaps take refuge in bio-empathy, and hope that something more solid one day turns up.

Many non-religious people would characterize their philosophical position as *humanist.* Humanism is not a very clear system of belief and takes numerous different forms, but a definition which I believe would apply to most people of this persuasion would be along the lines ‘a system of thought that rejects religious beliefs and centres on humans and their values, capacities, and worth’. Or, according to the American Humanist Association, the oldest and largest humanist society in the US:

> *Humanism is a progressive philosophy of life that, without supernaturalism, affirms our ability and responsibility to lead ethical lives of personal fulfilment that aspire to the greater good of humanity.*

The American Humanist Association issued a statement of principle – *Humanist Manifesto I* – in 1933, containing 15 propositions. None of these addressed environmental matters. Forty years later, however, the Association published *Humanist Manifesto II*, with 17 propositions. The 14th was:

> *The world community must engage in cooperative planning concerning the use of rapidly depleting resources. The planet earth must be considered a single ecosystem. Ecological damage, resource depletion, and excessive population growth must be checked by international concord. The cultivation and conservation of nature is a moral value; we should perceive ourselves as integral to the sources of our*

> *being in nature. We must free our world from needless pollution and waste, responsibly guarding and creating wealth, both natural and human. Exploitation of natural resources, uncurbed by social conscience, must end.*

What about Buddhism? The Buddhist world view is not theistic in that it does not hypothesize the existence of a creator of the universe, but at the same time it is not materialist atheist in that it assumes that there is another level of existence distinct from the physical universe. Can philosophically unattached conservationist, seeking a coherent basis for their beliefs, unwilling to take the theist road, find refuge in this ancient Asian religion? Maybe, but any conservationist who actually wants to achieve some practical outcomes in the way of preserving biodiversity may – as pointed out by zoologist Colin Tudge some years ago[30] – find the passivity and fatalism of Buddhism, compared to the hands-on, actively do-good, tradition of Christianity, something of a handicap.

CREATIONISM

Creationism is that belief system which holds not only that the universe was created by a non-material entity outside itself (the theist position) but that all past and present life-forms came into existence not by evolution from pre-existing forms, but as the result of a set of quite specific acts of creation. What does science have to say about this? Accepting that broadly speaking, setting aside the inevitable arguments about detail, palaeontologists have accurately characterized the life-forms preserved in the fossil record and that geologists have correctly determined their age and temporal sequence, then if we think that science tells us about objective reality we must conclude:

- that life on Earth is of great antiquity, with unicellular prokaryotes having existed for ~3.5 billion years and multicellular eukaryotes for at least 600 million years

- that an enormous variety of life-forms have existed in the past, which no longer exist
- that the complexity of life-forms has increased over time
- that within any given group (horses, whales, ferns) of organisms for which there is a significant fossil record there is evidence for sequential change, one type being followed by another.

The creationist hypothesis to explain all this cannot be disproved. An omnipotent being could do everything the creationists suppose. But why would God bother? Why go to the trouble of inserting into the geological record, by specific acts of creation at appropriate times in the Earth's history, a series of life-forms whose nature gives rise to the unavoidable impression that later forms arose naturally from earlier forms? A non-creationist theist might suppose that God could have taken the much more elegant approach of incorporating a built-in tendency towards evolutionary change, and the emergence of significant form, into the very fabric of physical reality. Einstein once said 'Subtle is the Lord but malicious he is not': to this we might add, that he is not whimsical either.

In short, a belief in creationism is not compatible with science.

Animal liberationism

In simple terms, animal liberationism holds that members of animal species should be accorded the same rights as members of the human species. It is deemed to follow from this that, because of the presumed accompanying suffering, we should not carry out animal experimentation, commercial farming or hunting of animals for food. Vegetarianism is considered to be the only ethically acceptable dietary practice. The contrary view to animal liberationism is pilloried as 'speciesism'. For rhetorical purposes speciesism is lumped together with sexism and racism, doubtless in the hope that some of the odium attached to those viewpoints will, without benefit of logic, somehow be transferred to people who presume to disagree with the animal liberationists.

The central proposition of animal liberationism, that animals should be given the same rights as humans, is simply an assertion: it is a statement of a particular emotional/psychological/moral position that some people in our society find attractive. It cannot be proven or disproved by logic. Science cannot provide any evidence for or against. Indeed, the proposition cannot meaningfully be described as true or untrue. Those who adopt this position do so because they feel good about it. In a free society they are fully entitled to do so, and it is in no way an aim of this book to argue that people should harbour only views that can be proven by mathematical logic or experimental evidence, since that would eliminate most of the emotional, aesthetic and other aspects of human experience which make life worth living.

Can science nevertheless bring anything to the debate? Perhaps a little, by clarifying some of the factual issues. The first point to be made is that the present body of medical knowledge was brought into being with, and could not have been achieved without, animal experimentation. Further advances in medical knowledge will also require animal experimentation. The same is true of veterinary research – carried out for the benefit of the very animals for which the liberationists express such concern. The view that perhaps we could do research from now on just using tissue culture, or perhaps even computer simulation, is delusory.

Although animal liberationism remains a minority position, there is a general consensus in the developed world that animals should be spared unnecessary suffering. This proposition assumes that animals, at least the higher mammals found in our laboratories or on our farms, are conscious beings, capable of suffering. As discussed in Chapter 4, they may not be. And if it ever became apparent that animals are not conscious, then we would have no need to worry about animal suffering. Nevertheless, to err on the safe side, let us assume for the time being that the mammalian species we use are indeed conscious and can suffer. Just how much suffering should be regarded as 'unnecessary' is hard to determine, since even the most pampered and protected human, let alone animal, life is

attended by some unavoidable suffering. As a possible benchmark we might say that particular practices in farming, hunting or experimenting upon animals should, on average, occasion no more suffering than those animal species would normally endure in the wild. How much suffering is that? Actually, rather a lot.

Ecology tells us that the wild relatives of our farm animals lead a very uncertain and dangerous life, chronically hungry, forever facing the prospect of murderous attack from the predators of whose food chain they form a part. Their end is always painful, and is frequently slow when age or chronic disease make them unable to outrun or outfly the carnivores who seek to eat them, or to fend off the scavengers and carrion eaters. Most of us have seen wildlife documentaries that show zebras being brought down by hungry lions, gazelles by leopards or moose by wolf packs, a rabbit in the curving talons of an eagle, a penguin in the jaws of a killer whale, a mouse in the owl's beak. Nature is indeed red in tooth and claw.

Contrast life in the wild with life experienced by animals on a well-run farm. Their food supply is guaranteed by the commercial imperative – the farmer will make a profit only if the animals produce meat, milk, wool or eggs at an economic rate. For the same reasons, their health is maintained: they are dosed against liver fluke and lungworm, dipped to prevent tick infestation, treated for blowfly strike and assisted through difficult pregnancies. They are protected from predators. The day of slaughter will of course come for every farm animal. Distressing though this prospect is for the tender-hearted, the plain fact is that every animal must die, and a very quick death in a modern abattoir allows the animal to experience only a small fraction of the suffering endured by its cousins in the wild. Furthermore, since farm animals have no foreknowledge of their impending death they do not suffer the anticipatory fear and foreboding that is the human lot. Interestingly, Peter Singer, one of the high priests of the animal liberation movement, has acknowledged that, given that an animal belongs to a species incapable of self-consciousness:

> *It is not wrong to rear and kill it for food, provided that it lives a pleasant life and, after being killed, will be replaced by another animal which will live a similarly pleasant life and would not have existed if the first animal had not been killed. This means that vegetarianism is not obligatory for those who can obtain meat from animals they know to have been reared in this manner.*[26]

In summary, properly managed modern farming can provide animals with an easy, well-fed and healthy life and a quick humane end. It can be argued that some forms of animal husbandry, such as egg production from battery hens, are inherently stressful for the animals. Where there is a consensus that this is the case, that form of production can be replaced by a more humane practice. Instead of battery hens, for example, we already have the alternative of egg production by birds on free range, or housed in barns. The essential point is that animal farming can be carried on without using inhumane methods of production.

What has been said about farm animals applies equally to laboratory animals used for experimentation. The animals are well-fed and housed. All animal experimentation in the developed world these days is governed by strict codes of practice designed to minimize stress and suffering. Traumatic procedures such as surgery are carried out under anaesthetic. Death, when it comes, is quick. The white laboratory rat, plump, warm and well-cared-for, has a much easier life than its lean grey cousins living on scraps in the urban wastelands, riddled with parasites and forever in danger of being eaten by the next hungry cat.

Animal liberationists, as we have seen, are opposed to farming of animals and to most forms of medical and veterinary research. What is not so generally realized is that they are also opposed to certain essential forms of ecosystem management. To manage ecosystems, especially those whose natural functioning has been distorted by human activities, it is frequently necessary to reduce the populations of particular animal species. These may be introduced animals preying upon or competing with the native fauna, or native herbivores which have multiplied beyond the ecosystem's

capacity to sustain them. Culling of such animals is carried out by environmental managers all around the world. To animal liberationists, however, this practice is anathema and they would rather see the ecosystem destroyed than have it conserved by such means.

A typical example occurred in 1986[1] when the US Fish and Wildlife Service sought to protect two endangered West Coast bird species, the light-footed clapper rail and the California least tern, in the Seal Beach Wildlife Refuge from non-native red foxes which had decimated the populations of both birds – down to five pairs in the case of the clapper rail. When biologists from the Service began trapping and killing the foxes, a group calling itself the 'Animal Lovers Volunteer Organization' sought to prevent the fox cull by court action. Fortunately for the endangered bird species, the animal liberationists failed.

A case in which the problem was the excessive multiplication of a native herbivore arose some years ago in the Hattah-Kulkyne National Park, Victoria, Australia. A lack of predators had allowed the kangaroo population to rise to the point at which, by overgrazing, it was causing substantial damage to the whole ecosystem. When the park managers began culling the kangaroos, to bring things back into balance and to ensure a viable habitat for kangaroos in the long term, animal liberationists physically attempted to prevent them.

To quote philosopher Elliott Sober:[27]

> *Both animal liberationists and environmentalists wish to broaden our ethical horizons – to make us realize that it is not just human welfare that counts. But they do this in very different, often conflicting, ways. It is no accident that at the level of practical politics the two points of view increasingly find themselves at loggerheads. This practical conflict is the expression of a deep theoretical divide.*

It is unfortunate that the popular media commonly confuse environmentalists and animal liberationists, since in reality the practicalities of environmental management and the ideology of animal liberation are unavoidably in conflict.

In summary, the contribution of science to the animal liberationism debate is that it tells us we can carry on both animal farming and animal research while still giving our animals a much more comfortable life than they would have in the wild. It also tells us that if the animal liberationist view on wild animal culling were to prevail, numerous ecosystems around the world would go into irreversible decline. In short, you can be an environmentalist or you can be an animal liberationist. You cannot be both.

FEMINISM

Does science have anything to say about feminism? This is a difficult question to answer since there are so many different versions of feminism. If we define feminism (and here I borrow from the entry[20] in *The Oxford Companion to Philosophy*) as a movement which 'attempts to attain equal legal and political rights for women', or in more extreme form as one 'which sees the relationship between the sexes as one of inequality, subordination, or oppression, and which aims to identify and remedy the sources of that oppression', then science, whatever personal views individual scientists may harbour, has nothing to say.

It is when feminists make certain specific claims about male–female differences that scientific considerations may become relevant. For example, it was quite common in the early days of the movement for feminists to assert that the psychological and behavioural differences between men and women are not innate but are culturally imposed by the ways in which boys and girls are brought up in our society. Partly in response to such claims, the nature and origin of male–female cognitive and emotional differences have now been extensively studied, female scientists taking a prominent role in this research. In a field as controversial as this, where so many protagonists have axes to grind, it may be doubted if we will ever be allowed to arrive at a consensus. For the time being, however, we can allow ourselves to be guided by the cautious and judicious contribution[17] by

Corinne Hutt and Miranda Hughes in *The Oxford Companion to the Mind*, which concludes *inter alia* that while cultural influences can act to persuade children to conform to sex-stereotyped expectations there are also biological sex differences which predispose males and females to some behavioural differences, and that there are sex differences in brain organization, notably the degree of lateralization of visuo-spatial and language skills between the hemispheres. Cambridge psychologist Simon Baron-Cohen summarizes male–female cognitive differences as revealed by published data as follows:[3]

> *Differences favouring males are seen on the mental rotation test, spatial navigation including map reading, targeting, and possibly the embedded figures test. Males are more likely to play with mechanical toys as children, and as adults score higher on engineering and physics problems.*
>
> *Females score higher on tests of emotion recognition, social sensitivity and verbal fluency. Girls start to talk earlier than boys do and are more likely to play with dolls as children.*

Any claim that male–female psychological differences are entirely culturally created is therefore contrary to what science seems to be telling us. We cannot, however, proceed to the conclusion that feminism at large is in error on this matter since within that movement a body of opinion has emerged which, far from denying it, proudly proclaims that woman's nature is different from man's and 'goes on to advocate a form of feminism which rejoices in that difference, and which argues for the revaluing of "women's qualities", qualities of emotion and intuition, above the "male" value of rationality'.[20]

Another area where there might be thought to be some interaction between feminism and science is that of 'feminist science'. Is there such a thing? The phrase is sometimes used by feminists in such a way as to imply that it is a meaningful category. My own view is that it is not. There is

science and there is non-science. Like the vast majority of practising scientists, I assume that there is a physical reality out there whose nature we seek to discover. The ideology of the investigator may well influence what they decide to work on, and it may sometimes dispose them to hope that one result rather than another will be found. There are certainly feminists engaged in scientific research but, given that they are competent and honest, reality must eventually prevail and, at least in those areas of science where clear conclusions can be reached, they will reach the same conclusions as their non-feminist colleagues. There are of course some areas of science, such as palaeo-anthropology, where the ratio of speculation and inference to actual fact is particularly high and there is the ever-present danger that ideology and culture may unconsciously influence interpretation. Nevertheless, the ideal of separating ideology from science is one that should be rigorously maintained. Categories such as 'feminist science', 'Marxist science' or any other '-ist science' should be regarded as unacceptable.

GENETIC DETERMINISM

Genetic determinism is that school of thought which claims that human beings are largely determined by our genes – not only physical attributes such as skin colour, height and blood group, but also most of our cognitive and behavioural characteristics such as aggression, conformism, gender-related roles, ethical beliefs, aesthetic tastes, intelligence and altruism. Genetic determinism is not a single precisely defined body of doctrine: its adherents cover a range from those who gleefully proclaim that we are just puppets blindly following the instructions laid down in our DNA, to those who believe that, while our behaviour and capabilities are very strongly influenced by our genes, we can nevertheless refuse to capitulate to our genetic predispositions.

For a long time this has been, and still is, a difficult and controversial area of science. Because of the enormously complex nature of humanity

and human societies, and the impossibility of carrying out the kinds of experiments on human beings that would be required, clear conclusions on matters of genes versus environment, nature versus nurture, are hard to come by. Although genetic determinist notions as applied to human beings were introduced to the public consciousness by pop-anthropological books such as Konrad Lorenz's *On Aggression*, Desmond Morris's *The Naked Ape* and Robert Ardrey's *The Territorial Imperative* in the 1960s, it did not really take off as a scientific controversy – one with very strong social-political-moral overtones – until the publication of *Sociobiology* by Edward O. Wilson in 1975. Although the book was almost entirely about the biological basis of social behaviour in animals, it included a final chapter – 'Man: from sociobiology to sociology' – in which Wilson speculated about the evolution, and thus by implication the genetic control, of culture, altruism, ethics, aesthetics, warfare, religion and other human attributes. In a later book, *On Human Nature*, Wilson asserted that 'The genes hold culture on a leash'.

Some scientists, particularly those concerned about unscrupulous people appropriating misunderstood science to further their own social and political agendas, felt great concern about *Sociobiology* and other Wilson books. Prominent members of this group of scientists were Richard Lewontin and Stephen J. Gould in the US and Stephen Rose in Britain (a detailed account of the whole controversy can be found in *Defenders of the Truth* by Ullica Segerstrale).[25] They feared that Wilson's pronouncements might be used to legitimize racism, and to justify the existing social and economic stratification of society on the grounds that people who have power, money and influence do so on the basis of their genetic superiority. For similar reasons there was concern about the genetic triumphalism of Richard Dawkins' book *The Selfish Gene*. This sets out a gene's-eye view of biology and, although for the most part was concerned with animals rather than people, it characterized human beings as 'survival machines – robot vehicles blindly programmed to preserve the selfish molecules known as genes'. The perceived danger was that acceptance of such genetic

determinism might be used to exonerate those responsible for war, racist oppression and so on. The fears of the concerned scientists were not without basis. The far-right wing National Front in Britain, for example, cited Wilson and Dawkins in support of its proposition that nationalism and racism are products of our genes (and thus, by implication, natural and proper views to espouse).

In fact neither Wilson nor Dawkins supported these extreme views. On the question of genetically based social stratification, for example, Wilson's last chapter of *Sociobiology* said:

> *Human societies have effloresced to levels of extreme complexity because their members have the intelligence and flexibility to play roles of virtually any degree of specification, and to switch them as the occasion demands.*

And again:

> *Yet despite the plausibility of the general argument, there is little evidence of any hereditary solidification of status … Powerful forces can be identified that work against the genetic fixation of caste differences … A much more likely circumstance is the one that apparently prevails: the maintenance of a large amount of genetic diversity within societies and the loose correlation of some of the genetically determined traits with success.*

In addition, in a letter to *Nature*[33] explicitly dissociating himself from the National Front attempt to use sociobiology to further its cause, Wilson said that there is 'no justification for racism to be found in the truly scientific study of the biological basis of social behaviour'. While acknowledging that there might be a hereditary tendency to acquire xenophobic and nationalist feelings, he argued that:

> *It is a* non sequitur *to interpret such a hypothesis as an argument in favour of racist ideology. It is more reasonable to assume that a knowledge of such a hereditary basis can lead to circumvention of destructive behaviour such as racism.*

Dawkins also dissociated himself from the implications that the National Front sought to derive from the 'selfish gene' concept, saying in a letter to Nature:[12]

> *What is really wrong with the National Front quotation is not the suggestion that natural selection favoured the evolution of a tendency to be selfish and even racist. What I object to is the suggestion that if such tendencies had evolved they would be inevitable and ineradicable; the suggestion that we are stuck with our biological nature and can't change it.*

In addition, despite the general portrayal in *The Selfish Gene* of living creatures as automata, behaving in the manner specified by the algorithms written in their DNA, Dawkins ends his book with a rousing call to arms: 'We alone on Earth can rebel against the tyranny of the selfish replicators.'

The controversy over the extent to which the capabilities and behaviour patterns of human beings are determined by their genes rather than their environment is by no means over. The clamour of battle, however, has diminished somewhat and the dust shows signs of settling. What can science now contribute with any confidence to this debate? Not a great deal, because experimental resolution of these issues, such as we might implement with animals, will always be out of the question. The only evidence comes from observation of human society and human individuals as we find them, with the consequent impossibility of controlling all the relevant variables.

For comparison purposes, consider how a plant breeder would compare the performance of two different varieties of a crop plant such as wheat, canola or soybean. The varieties will be planted side-by-side in land carefully selected for uniformity, and suitability for the crop in question. Several replicated plots of each variety will be sown, and each plot will contain hundreds or perhaps thousands of individual plants. Care will be taken to ensure that all the plots receive exactly the same treatment in terms of soil preparation, irrigation and spray regime. After harvest, the

results will be analysed to see if statistically significant differences in yield can be discerned. If the experimenter learns at some stage that an important variable had not been adequately controlled, e.g. that owing to equipment malfunction some plots were irrigated more than others, they do not attempt to fiddle with the data in the hope of correcting for this difference: the results go into the wastepaper basket and the experiment is carried out again.

For would-be human-behavioural geneticists, such rigour is an impossible dream. They have to seek their experimental material where they can find it, and must reconcile themselves to the fact that their chosen samples of the groups that they wish to compare for a particular character will never have been reared in identical environments. The only hope is that accumulating enough data will ultimately begin to cancel out the differences, and that a faint signal may eventually be discerned through the fog of statistical noise.

The least-equivocal evidence comes from comparative studies of identical and non-identical twins. Identical twins have the same genetic makeup; non-identical twins share only half their genes. The two members of any pair of twins are in most cases reared in essentially the same home environment. Thus, the observation that a particular behavioural characteristic is more strongly correlated in identical than in non-identical twins is evidence for a genetic contribution. An interesting example is a study by a joint US–British team[13] of musical pitch perception in 136 monozygotic (identical) and 148 dizygotic (non-identical) twin pairs. The experimenters measured the ability of their subjects to recognize whether simple popular melodies contain notes with incorrect pitch. The correlation in performance was substantially higher (0.67) in the monozygotic than in the dizygotic (0.44) twins. On the basis of the data the experimenters estimated that the heritability of musical pitch recognition in humans (the extent to which it is determined by genetic factors) is 71–80%.

Comparison of identical twins who have been reared apart in different home environments by mischance can – by identifying which features of

personality and ability remain the same, and which tend to change – provide pointers to which characters are genetically determined. In the nature of things, however, there is not a lot of this sort of data in existence.

To say that human-behavioural genetics is a field in which few clear generalizations can yet be made is not to imply that human temperament and cognitive abilities are entirely environmentally determined, and to no extent under genetic control. Observation tells us, for example, that within some higher animal species aggressive tendencies can be inherited, as evidenced by Spanish fighting bulls and pitbull terriers. We must recognize the possibility that within the human population some people are genetically more predisposed than others towards aggression. As far as ability is concerned there is evidence – some coming from twin studies – for a strong genetic contribution to intelligence, at least as measured by IQ tests. We all know from our own observation that intelligence is highly variable within the human population, and the evidence to date indicates that the contribution of heredity to variance (a statistical measure of variability) is about 80%. This applies to the variance in IQ *within* a group, e.g. white Caucasians in North America. Problems arise when people seek to apply the same genetic explanation to measured differences of average IQ *between* groups, e.g. to the fact that the average IQ score of African Americans in the US is somewhat lower than that of Caucasians. Given the great effect of environmental influences on mental development, and also given the marked disparity between social and economic situations of different groups in the US, my own view is that to attribute the difference in average IQ to a genetic cause is not justified on the evidence currently available. My scepticism is increased by the fact that the measured average IQ of human populations is now observed to vary in time for reasons that are not understood. James Flynn, a political scientist working in New Zealand, has analysed IQ measurement data published in the literature over several decades and shown that average IQ for 20 or so countries for which data is available has increased by about 10 points (out of a notional previous average of 100) over a generation. No-one knows why. In the same

time the gap between African and Caucasian Americans has decreased by about 5 points.[18] Trying to establish a genetic basis for observed differences between ethnic groups in such a variable, environmentally sensitive and ill-understood parameter as IQ is, I suggest, an enterprise doomed to failure.

Nevertheless, regardless of whether we can establish real genetically-based differences between groups, the differences in the various kinds of intelligence (musical, mathematical, verbal, visual-artistic) and in temperament (placid, excitable, pacific, aggressive, intellectual, pragmatic, sociable, reclusive) between individuals must surely be partly genetically determined – they are partly the product of our detailed brain structure and hormonal makeup. This does not mean that human beings run on genetic railway lines from which they cannot diverge. We have enormous versatility and learning capacity as well as our genetic heritage, and so although each of us may have genetically determined limits of capability these can be wide if we are prepared to make the effort. While we may have certain temperamental propensities, we do not have to give in to them. Given time and good teaching, the mathematics-averse can grasp the mathematical essentials. The aggressive person can strive to control their anger. The recluse can make the effort occasionally to be sociable.

Some writers of a genetic determinist persuasion go much further than I have here in the influence of genetics in human affairs. The late C.D. Darlington, eminent cytogeneticist and Professor of Botany at Oxford (in whose department I was a post-doctoral fellow many years ago), wrote *The Evolution of Man and Society* which presented a picture of the whole of human history refracted through a genetic prism. He interpreted the succession of historical events, the ebb and flow of cultures, very much in terms of the interaction and competition between a variety of castes and tribes variously exhibiting this or that genetically endowed ability. He attached – speculatively, but plausibly – particular importance to the necessity for outbreeding in order to maintain adaptability and vigour within any group, such as a ruling elite, and provided many examples of the disastrous consequences of inbreeding.

Zoologist and sociobiologist Edward Wilson claims that 'the genes hold culture on a leash'. A contrary view which I find much more persuasive is that when self-reflective consciousness emerged in *Homo sapiens* ~150 000 years ago, it cut the leash. Human culture emerged and proliferated wildly in all kinds of unlikely directions which could never have been predicted on the basis of genetics. Genetic influences on human behaviour and culture undoubtedly exist but they are better likened to a gentle wind than to a leash: a wind that we can sail at an angle to, or even against, if we wish.

References and further reading

Chapter 1

1 Barrow, John D. & Tipler, Frank, J. (1996). *The Anthropic Cosmological Principle.* Oxford University Press: Oxford.

2 Bennett, Charles L. (2006). Cosmology from start to finish. *Nature* **440**, 1126–1131.

3 Branch, David (1989). Supernovae. Cited in Ellis, George F.R. (2003). The shape of the universe. *Nature* **425**, 733–751.

4 Carroll, Sean M. (2006). Is our universe natural? *Nature* **440**, 1132–1136.

5 Chambers, John E. (2004). Planetary accretion in the inner solar system. *Earth and Planetary Science Letters* **223**, 241–252.

6 Cowan, John J. & Sneden, Christopher (2006). Heavy element synthesis in the oldest star and the early universe. *Nature* **440**, 1151–1156.

7 Craig, William Lane & Smith, Quentin (1993). *Theism, Atheism and Big Bang Cosmology.* Clarendon Press: Oxford.

8 Davies, Paul (1995). *Superforce.* 2nd edn. Penguin Books: London.

9 Delsemme, Armand (1998). *Our Cosmic Origins.* Cambridge University Press: Cambridge.

10 Ellis, George F.R. (2003). The shape of the universe. *Nature* **425**, 566–567.

11 *Encyclopedia of Astronomy and Astrophysics.* (1989). Robert A. Meyer (ed.). Academic Press: San Diego.

12 Fukugita, Masataka (2003). The dark side. *Nature* **422**, 489–491.

13 Gribbin, John (1998). *In Search of the Big Bang.* 2nd edn. Penguin Books: London.

14 Guth, Alan (1998). *The Inflationary Universe.* Vintage: London.

15 Hu, Wayne & White, Martin (2004). The cosmic symphony. *Scientific American* **290**, 44–53.

16 Larson, Richard B. & Bromm, Volker (2004). The first stars in the universe. *Scientific American Special Edition* **14**, 4–11.

17 Luminet, Jean-Pierre, Weeks, Jeffrey R., Riazuelo, Alain, Lehoucq, Roland & Uzan, Jean-Phillippe (2003). Dodecahedral space topology as an explanation for weak wide-angle temperature correlations in the cosmic microwave background. *Nature* **425**, 593–595.

18 NASA WMAP website. map.gsfc.nasa.gov.

19 Starkman, Glenn D. & Schwarz, Dominik J. (2005). *Scientific American* **293**, 36–43.

20 Singh, Simon (2005). *Big Bang.* Harper Perennnial: London.

21 Tryon, Edward P. (1973). Is the universe a quantum fluctuation? *Nature* **246**, 396–397.

22 Weinberg, Stephen (1977). *The First Three Minutes: A Modern View of the Origin of the Universe.* Basic Books: New York.

23 Weisskopf, Victor F. (1983). The origin of the universe. *American Scientist* **71,** 473–480.

Chapter 2

1 Allwood, A.C., Walter, M.R., Kamber, B.S., Marshall, C.P. & Burch, I.W. (2006). Stromatolite reef from the Early Archaean era of Australia. *Nature* **441**, 714–718.

2 Awramik, S.M. (2006). Respect for stromatolites. *Nature* **441**, 700–701.

3 Britten, Roy J. (2002).Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proceedings of the National Academy of Sciences* (U.S.) **99**, 13633–13635.

4 Brocks, J.J., Logan, G.A., Buick, R. & Summons, R.E. (1999). Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036.

5 Brocks, J.J., Buick, R., Summons, R.E. & Logan, G.A. (2003). A reconstruction of Archaean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Roger Supergroup, Hamersley Basin, Western Australia. *Geochimica et Cosmochimica Acta* **67**, 4321–4335.

6 Carroll, Sean, B. (2005). *Endless Forms Most Beautiful: the New Science of Evo Devo.* W.W. Norton: New York.

7 Clack, J.A. (2005). Getting a leg up on land. *Scientific American* **293**, 80–87.

8 Conway-Morris, Simon (2003). *Life's Solution: Inevitable Humans in a Lonely Universe.* Cambridge University Press: Cambridge.

9 Cooper, G., Kimmich, N., Belisle, W., Sarinana, J., Brabham, K. & Garrel, L. (2001). Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth. *Nature* **414**, 879–883.

10 Dawkins, Richard (2005). *The Ancestor's Tale: A Pilgrimage to the Dawn of Life.* Phoenix: London.

11 Delsemme, Armand (1998). *Our Cosmic Origins.* Cambridge University Press: Cambridge.

12 *Encyclopaedia Britannica.* Sections on Man, evolution of; Primates; Evolution, human; Paleontology; etc. Any modern edition. Encyclopaedia Britannica Inc.:Chicago.

13 Fortey, Richard (1998). *Life: An Unauthorised Biography.* Flamingo: London.

14 Goldsmith, T.H. (2006). What birds see. *Scientific American* **295**, 51–57.

15 Goodwin, Brian (1994). *How the Leopard Changed its Spots.* Scribner: New York.

16 Gould, Stephen Jay (1989). *Wonderful Life: The Burgess Shale and the Nature of History.* W.W. Norton & Co.: New York.

17 Javaux, E.J., Knoll, A.H. & Walter, M.R. (2004). TEM evidence for eukaryotic diversity in mid-Proterozoic oceans. *Geobiology* **2**, 121–132.

18 Joyce, Gerald F. (2002). The antiquity of RNA-based evolution *Nature* **418**, 214–221.

19 Kasting, J.F., Pavlov, A.A. & Siefert, J.L. (2001). A coupled ecosystem-climate model for predicting the methane concentration in the Archean atmosphere. *Origins of Life and Evolution of Biospheres* **31**, 271–285.

20 Kauffman, Stuart (1995). *At Home in the Universe.* Viking: London.

21 Knoll, A.H. (1992). The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627.

22 Knoll, Andrew, H. (2003). *Life on a Young Planet: The First Three Billion Years of Evolution on Earth.* Princeton University Press: Princeton.

23 Leakey, Richard & Lewin, Roger (1993). *Origins Reconsidered.* Abacus: London.

24 Margulis, Lynn (1993). *Symbiosis in Cell Evolution.* Freeman: San Francisco.

25 Martin, W. & Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41.

26 Martin, W. & Russell, M.J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society, London, B* **358**, 59–85.

27 Moffatt, A.S. (2002). New fossils and a glimpse of evolution. *Science* **295**, 612–613.

28 Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P. & Friend, C.R. (1996). Evidence for life on Earth more than 3800 million years ago. *Nature* **384**, 55–59.

29 Narbonne, G.M. & Gehling, J.G. (2003). Life after snowball: the oldest complex Ediacaran fossils. *Geology* **31**, 27–30.

30 Pavlov, A.A., Hurtgen, M.T., Kasting, J.F. & Arthur, M.A. (2003). Methane-rich Proterozoic atmosphere. *Geology* **31**, 87–90.

31 Russell, M. (2006). First life. *American Scientist* **94**, 32–39.

32 Russell, M.J. & Hall, A. (1997). The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *Journal of the Geological Society, London* **154**, 377–402.

33 Rye, R. & Holland, H.D. (2000). Life associated with a 2.76 Ga ephemeral pond?: evidence from Mount Roe #2 paleosol. *Geology* **28**, 483–486.

34 Sephton, M.A. (2001). Life's sweet beginnings. *Nature* **414**, 857–858.

35 Stubbs, Lisa. (2006). Human Genome Project Information website. www.ornl.gov./sci/techresources/Human_Genome/faq/compgen.shtml.

36 Tan, Y. & Li, W.-H. (1999). Trichromatic vision in prosimians. *Nature* **402**, 36.

37 Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. (2006). Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519.

38 Wächtershäuser, G. (1988). Pyrite formation, the first energy source for life: a hypothesis. *Systematic & Applied Microbiology* **21**, 473–477.

Chapter 3

1 Arnt, M., Hairz, O., Vos-Andreae, J., Keller, C., van der Zouw, G. & Zeilinger, A. (1999). Wave-particle duality of C_{60} molecules. *Nature* **401**, 680–682.

2 Baggott, Jim (1992). *The Meaning of Quantum Theory.* Oxford University Press: Oxford.

3 Bell, John S. (1987). *Speakable and Unspeakable in Quantum Mechanics.* Cambridge University Press: Cambridge.

4 Berry, Michael (1988). In Wolpert, L. & Richards, A. (eds). *A Passion for Science.* Oxford University Press: Oxford, pp. 39–51.

5 Brumfiel, G. (2006). Outrageous fortune. *Nature* **439**, 10–12.

6 Davies, P.C.W. (1984). *Quantum Mechanics.* Routledge & Kegan Paul: London.

7 Davies, Paul (1990). Chaos frees the universe. *New Scientist* 6 October 1990.

8 Davies, Paul (1995). *Superforce.* 2nd edn. Penguin Books: London, pp. 104–105.

9 Davies, Paul & Gribbin, John (1992). *The Matter Myth.* Penguin Books: London.

10 Dyson, Freeman (2004). The World on a string. *New York Review of Books* **LI: 8** (13 May), 16–19.

11 Hey, Tony & Walters, Patrick (1987). *The Quantum Universe.* Cambridge University Press: Cambridge.

12 Jammer, Max (1989). *The Conceptual Development of Quantum Mechanics.* 2nd edn. American Institute of Physics.

13 Joos, Erich (1999). In Blanchard, P., Giulini, D., Joos, E., Kiefer, C. & Stamatescu, I.-O. (eds). *Decoherence: Theoretical, Experimental and Conceptual Problems.* Springer: Berlin.

14 Penrose, Roger (1994). *Shadows of the Mind.* Oxford University Press: Oxford, pp. 237–339.

15 Penrose, Roger (1997). *The Large, the Small and the Human Mind.* Cambridge University Press: Cambridge, ch. 2.

16 Schlosshauer, Maximilian (2004). Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews in Modern Physics* **76**, 1267–1305.

17 Selleri, Franco (1989). *Quantum Paradoxes and Physical Reality.* Kluwer: Dordrecht.

18 Squires, Euan (1986). *The Mystery of the Quantum World.* Adam Hilger: Bristol.

19 Squires, Euan (1990). *Conscious Mind in the Physical World.* Adam Hilger: Bristol.

20 Tegmark, Max & Wheeler, John (2001). 100 years of quantum mysteries. *Scientific American* **284**, 68–75.

Chapter 4

1 Beck, F. & Eccles, J.C. (1992). Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Science*, USA **89**, 11357–11361.

2 Chalmers, David J. (1996). *The Conscious Mind.* Oxford University Press: New York.

3 Eccles, John C. (1994). *How the Self Controls its Brain.* Springer: Berlin.

4 Fodor, J.A. (1992). The big idea. Can there be a science of mind? *Times Literary Supplement* 3 July, 5–7.

5 Gregory, Richard L. (ed.) (1987). T*he Oxford Companion to the Mind.* Oxford University Press: Oxford.

6 Hagan, S., Hameroff, S.R. & Tuszynski, J.A. (2000). Quantum computation in brain microtubules: decoherence and biological feasibility. In *Towards a Science of Consciousness.* Conference, Tucson Arizona, April 2000.

7 Honderich, Ted (ed.) (1995). *The Oxford Companion to Philosophy.* Oxford University Press: Oxford.

8 Penrose, Roger (1989). *The Emperor's New Mind.* Oxford University Press: New York.

9 Penrose, Roger (1994). *Shadows of the Mind.* Oxford University Press: Oxford.

10 Penrose, Roger (1999). *The Large, the Small and the Human Mind.* Cambridge University Press: Cambridge.

11 Popper, Karl R. & Eccles, John C. (1977). *The Self and its Brain.* Springer: Berlin.

12 Redman, S.J. (1990). Quantal analysis of synaptic potentials in neurons of the central nervous system.*Physiological Reviews* **70**, 165–198.

13 Rose, Stephen (ed.) (1998). *From Brains to Consciousness?* Allen Lane/Penguin Press: London.

14 Squires, Euan (1990). *Conscious Mind in the Physical World.* Adam Hilger: Bristol.

15 Stapp, Henry P. (1993). *Mind, Matter and Quantum Mechanics.* Springer: Berlin.

16 Tegmark, Max (2000). Importance of quantum decoherence in brain processes. *Physical Review E* **61**, 4194–4206.

17 Wynne, C.D.L. (2004). *Do Animals Think?* Princeton University Press: Princeton.

18 Wynne, C.D.L. (2004). The perils of anthropomorphism. *Nature* **428**, 606.

Chapter 5

1 Bramwell, Anna (1989). *Ecology in the 20th Century: A History.* Yale University Press: New Haven.

2 Devall, Bill & Sessions, George (1985). *Deep Ecology.* Peregrine Smith: Salt Lake City.

3 Flannery, Tim (1994). *The Future Eaters: An Ecological History of the Australasian Lands and People.* Reed Books: Melbourne.

4 Goldsmith, Edward. (2002). How can we survive? *The Ecologist* **32**, 24–27.

5 Grove, Richard (1990). The origins of environmentalism. *Nature* **345**, 11–14.

6 Grove, Richard (1992). Origins of western environmentalism. *Scientific American* July, 22–27.

7 Hay, Peter (2002). *Main Currents in Western Environmental Thought.* UNSW Press: Sydney.

8 Leopold, Aldo (1968). *A Sand County Almanac.* Oxford: New York.

9 Ponting, Clive (1993). *A Green History of the World: The Environment and the Collapse of Great Civilizations.* Penguin Books: New York.

10 Yi-Fu Tuan (1970). Our treatment of the environment in ideal and actuality. *American Scientist* **58**, 244–249.

Chapter 6

1 Anderson, Luke (2000). *Genetic Engineering, Food and our Environment.* Scribe Publications: Melbourne.

2 Falkow, Stanley & Kennedy, Donald (2001). Antibiotics, animals, and people – again! *Science* **291**, 19 January, editorial.

3 Flannery, Tim (1994). *The Future Eaters: An Ecological History of the Australasian Lands and People.* Reed Books: Melbourne.

4 Flannery, Tim (2005). *The Weather Makers.* Text Publishing: Melbourne.

5 Grove, Richard H. (1996). *Green Imperialism: Colonial Expansion, Tropical Island Edens and the Origins of Environmentalism.* Cambridge University Press: Cambridge.

6 Hannum, W.H., Marsh, G.E. & Stanford, G.S. (2005). Smarter use of nuclear waste. *Scientific American.* **294**, 84–91.

7 Kaiser, J. (2002). Livestock feed ban preserves drugs' power. *Science.* **295**, 28-29.

8 Lappé, Marc & Bailey, Britt (1999). *Against the Grain: The Genetic Transformation of Global Agriculture.* Earthscan Publications: London.

9 Loewen, E.P. (2004). Heavy-metal nuclear power. *American Scientist.* **92**, 522-531.

10 Lutz, W., O'Neill, B.C. & Scherbov, S. (2003). Europe's population at a turning point. *Science.* **299**, 19-20.

11 National Statistics (U.K.) News Release, September 2004.

12 Ponting, Clive (1993). *A Green History of the World: The Environment and the Collapse of Great Civilizations.* Penguin Books: New York.

13 Powell, K. (2006). Agricultural environmental and industrial biotechnology. *Nature Biotechnology.* **24**, 294.

14 Schrope, Mark (2001). Biologists urge US to build marine reserves. *Nature* **409**, 971.

15 US Department of Energy (2005). *Basic Research Needs for Solar Energy Ultilization.* DOE: Washington.

16 US Department of Energy (2006). *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda.* DOE: Washington.

17 Wilson, Edward O. (1992). *The Diversity of Life.* Harvard University Press: Cambridge, Mass.

Chapter 7

1 Archer, S.D., Stelfox-Widdicombe, C.E., Malin, G. & Burkill, P. (2003). Is dimethyl sulphide production related to microzooplankton herbivory in the southern North Sea? *Journal of Plankton Research* **25,** 235–242.

2 Charlson, R.J., Lovelock, J.E., Andreae, M.O. & Warren, S.G. (1987). Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**, 655–661.

3 Lovelock, James (1979). *Gaia – A New Look at Life on Earth.* Oxford University Press: Oxford.

4 Lovelock, James (1988). *The Ages of Gaia.* Oxford: New York.

5 Lovelock, James (1990). Hands up for the Gaia hypothesis. *Nature* **344**, 100–102.

6 Lovelock, James. (2003). The living Earth. *Nature* **426**, 769–770.

7 Moore, Walter J. (1956). *Physical Chemistry.* Longman: London.

8 O'Dowd, C.D., Jimenez, J.L., Bahreini, R., Flagan, R.C., Seinfeld, J.H., Hämeri, K., Pirjola, L., Kulmala, M., Jennings, S.G. & Hoffman, T. (2002). Marine aerosol formation from biogenic iodine emissions. *Nature* **417**, 632–635.

9 Sagan, Dorion & Margulis, Lynn (1983). The Gaian perspective of ecology. *The Ecologist* **13**, 160–167.

10 Schneider, Stephen (1990). Debating Gaia. *Environment* **32**, 5–32.

11 Schwartz, S.E. (1988). Are global cloud albedo and climate controlled by marine phytoplankton? *Nature* **336**, 441–445.

12 Schwartzman, David W. & Volk, Tyler (1989). Biotic enhancement of weathering and the habitability of Earth. *Nature* **340**, 457–460.

13 Watson, Andrew (1991). Gaia. *New Scientist* 6 July, 1–4.

Chapter 8

1 Alper, Joe (1991). Outfoxing the foxes. *Science* **253**, 741.

2 Barrow, John D. & Tipler, Frank J. (1996). *The Anthropic Cosmological Principle.* Oxford University Press: Oxford.

3 Baron-Cohen, S., Knickmeyer, R.C. & Belmonte, M.K. (2005). Sex differences within the brain: implications for explaining autism. *Science* **310**, 819–823.

4 Brumfiel, G. (2006). Outrageous fortune. *Nature* **439**, 10–12.

5 Callicott, J. Baird (1986). On the intrinsic value of nonhuman species. In Norton, Bryan G. (ed.). *The Preservation of Species.* Princeton University Press: Princeton, pp. 138–172.

6 Carter, Brandon, quoted by Davies, Paul (1995). *Superforce.* Penguin Books: London, p. 242.

7 Christiansen, Drew & Grazer, Walter (1996). *'And God Saw That It Was Good': Catholic Theology and the Environment.* United States Catholic Conference: Washington, DC.

8 Conway Morris, Simon. (2006). In review of Wilson, E.O. *The Creation: An Appeal to Save Life on Earth. Nature* **443**, 273.

9 Darlington, C.D. (1969). *The Evolution of Man and Society*. George Allen & Unwin: London.

10 Davies, Paul (1995). *Superforce*. Penguin Books: London.

11 Dawkins, Richard (1976). *The Selfish Gene*. Oxford University Press: New York.

12 Dawkins, Richard (1981). Selfish genes in race or politics. *Nature* **289**, 528.

13 Drayna, D., Manichaikul, A., de Lange, M., Snieder, H. & Spector, T. (2001). Genetic correlates of musical pitch perception in humans. *Science* 1969–1972.

14 Dyson, Freeman (1981). *Disturbing the Universe*. Pan Books: London.

15 Hay, Peter (2002). *Main Currents in Western Environmental Thought*. UNSW Press: Sydney.

16 Hilbert, David. Quoted by Ellis, George (2005). In review of Susskind, L. (2005). *Nature* **438**, 739–740.

17 Hutt, Corinne & Hughes, Miranda (1987). Sex differences in childhood. In Gregory, Richard L. (ed.). *The Oxford Companion to the Mind*. Oxford University Press: Oxford, pp. 701–703.

18 Jones, Steve (1996). *In the Blood: God, Genes and Destiny*. HarperCollins: London.

19 McDonagh, Sean (1986). *To Care for the Earth: A Call to a New Theology*. Geoffrey Chapman: London.

20 Mendus, Susan (1995). Feminism. In Honderich, Ted (ed.). *The Oxford Companion to Philosophy*. Oxford University Press: Oxford, pp. 270–272.

21 Polkinghorne, John (1998). *Belief in God in an Age of Science*. Yale University Press: New Haven.

22 Pope John Paul II (1990). The ecological crisis: a common responsibility. Message for Celebration of World Day of Peace. In Christiansen, Drew & Grazer, Walter (1996). *'And God Saw That It Was Good': Catholic Theology and the Environment*. United States Catholic Conference: Washington, DC, pp. 215–222.

22a Pope John Paul II (2000). We must be the custodians of nature. *L'Osservatore Romano* (weekly edition in English). 15 November, p. 1.

23 Rees, Martin (2001). *Our Cosmic Habitat*. Weidenfeld & Nicholson: London.

24 Rolston, Holmes III (1999). *Genes, Genesis and God: Values and Their Origins in Natural and Human History*. Cambridge University Press: Cambridge.

25 Segerstrale, Ullica (2000). *Defenders of the Truth: The Battle for Science in the Sociobiology Debate and Beyond*. Oxford University Press: Oxford.

26 Singer, Peter (1979). Killing humans and killing animals. *Inquiry* **22**, 145–156. Quoted in Hay, Peter (2002). *Main Currents in Western Environmental Thought*. UNSW Press: Sydney, p. 40.

27 Sober, Elliott (1986). Philosophical problems for environmentalism. In Norton, Bryan G. (ed.). *The Preservation of Species*. Princeton University Press: Princeton, pp. 173–194.

28 Susskind, Leonard (2005). *The Cosmic Landscape: String Theory and the Illusion of Cosmic Design*. Little Brown: New York.

29 The Evangelical Climate Initiative (2006). *Climate Change: An Evangelical Call to Action.* See also www.christiansandclimate.org.

30 Tudge, Colin (1991). *New Scientist* 13 April, 4.

31 United States Catholic Bishops (1991). Pastoral statement – renewing the earth. In Christiansen, Drew & Grazer, Walter (1996). *'And God Saw That It Was Good': Catholic Theology and the Environment.* United States Catholic Conference: Washington, DC, pp. 223–243.

32 Wilson, Edward O. (1975). *Sociobiology: The New Synthesis.* Harvard University Press: Cambridge.

33 Wilson, Edward O. (1981). Genes and racism. *Nature* **289**, 627.

34 Wilson, Edward O. (1992). *The Diversity of Life.* Harvard University Press: Cambridge.

INDEX

Acanthostega 59
accretion disk 21–23
Acetabularia 78
action potential 129–30
adenine 37
African Americans 239–40
aggression 239
Agrobacterium tumefaciens 194
Akilia Island 39
algae 56, 178, 205
algorithms 122–23, 237
alien species 175–76, 180
aliphatic carboxylic acids 36
Allwood, Abigail 3
alpha particles 86–91, 95
Alvarez, Luis 65
Alvarez, Walter 65
amino acids 36–38
ammonia 20, 36, 38, 45, 47
amphibians 59–60, 62, 75
angular power spectrum 28
animal culling 230–32
animal experimentation 228, 230
animal liberationism 227–32
Antarctica 162, 191, 193
anthracosaurs 60, 62
anthropic cosmological principle 5, 11, 215
anthropocentrism 149, 155
anthroposophy 150
antibiotics 76, 184–88
  in livestock rearing 186–8
  overprescription 186
antimatter 7–9
antineutrinos 7, 9, 31–32
antineutron 25
antiquarks 7
apes 61, 75
  human-like 71
  tailless 69–70
aquaculture 183
archaea 44
archaebacteria 44, 54–55
Archean eon 35
Ardipithecus ramidus 70
Ardrey, Robert 235
arthropods 58, 62
Aspect, Alain 99–100
asteroid 65–68
atheism 213–26
  and environmentalism 221–26
atmosphere 34, 36, 50–53
Australia 170, 172, 175, 176, 177, 180–81
Australopithecus 61, 71
axon 126–30

Bacillus thuringiensis 195
backbones 59
bacteria 43–45, 204
  antibiotic-resistant 184–88
  gene transfer 185–86
Baron-Cohen, Simon 233
Barrow, John 5, 215
baryons 25, 27
battery hens 230
Beck, Friedrich 133
behaviourism 139
Bell, John 99
Berkeley, George 117
Berry, Michael 108–9
beryllium 17, 216
beta decay 31
Big Bang 3, 5–6, 10, 13, 27, 79
biocentric philosophy 151, 156–57
biodiesel 165
biodiversity 171–73, 175–77, 179–81
biodynamic agriculture 159
bio-empathy 223, 225
biofuels 164–66
biological control 180–81
biological pump 202–04
biological resources 173–77, 179
biomass 202

biomass energy 165
biosphere 56, 155, 201, 204–06, 210–11
Birch, Charles 132
birds 64, 141
  species extinction 172
black body 3, 84
Bohm, David 106–7
Bohr, Niels 81, 93
Bookchin, Murray 156
brain 126
  and consciousness 125–26, 135, 220–21
  avian 141
  structures within 126
  volume 71–72
Bramwell, Anna 148–149
bromide 36
bromine 192
Brunet, Michel 70
Buddhism 226
Byron, Lord George Gordon 143

Cactoblastis insect 180
calcium carbonate 202, 21
California least tern 231
Callicott, J. Baird 222–23
Cambrian Explosion 57
Cambrian period 35, 57, 59–60
Campylobacter 187
canola 164–65, 169, 237
carbides 20
carbohydrate 47, 165, 204
carbon 15–20, 24, 31, 39
  formation 17, 216–17
carbon dioxide 20, 36, 38, 39, 45, 148
  reduction to carbohydrate 46–50, 204
  and atmospheric warming 53, 160, 162–66
  in present-day atmosphere 162–63
  regulation of atmospheric 200–05, 207–08
  silicate weathering 207, 211
carbon isotopes 38
carbon monoxide 20, 38
carbonic acid 148, 207, 211
Carboniferous period 35, 60, 62
carnivores 229
Carson, Rachel 152
Carter, Brandon 216
cataracts 192
Catholics 224–25
Caucasians 239–40
causality 104
cellulose 165
Cenozoic era 35
cerebral cortex 69, 126, 139, 141
Chad 70
chaos (indeterminism) 107–10
Charlson, Robert 203
chemoautotrophes 44–46
Chernobyl 166, 189, 195
chert 39
Chicxulub 65
chimpanzee 70, 73–4
chloride 36
chlorine 191–93
chlorofluorocarbons 159
  effect on ozone layer 191–93, 197
chloroplasts 43, 45, 53
  origin 55–56
chordates 58, 75
Christianity 224–26
chromosomes 42, 54, 194
Clack, Jennifer 59
clapper rail 231
cloud formation 203, 208
comets 23–24, 36–37
  and K-T boundary 65–6
Commoner, Barry 152
computer modelling 108
computers 121–25
conifers 69
conscious observer 101–3
consciousness 115–42, 220–21, 241
  and expression of emotion 138–39
  and quantum mechanics 133–38
  and sense impressions 116–18
  and the brain 125–30
  in animals 118, 138–42, 228
  in evolution 141–42
  metaphysics of 130–32
corals 179
Cosmic Dark Age 11, 14

cosmic microwave background 3, 11–12, 26–28
cosmological constant 27, 215
cosmologists 1, 2, 5, 27, 218
Coulomb, Charles Augustin 83
craters 23–24
Creationism 226–27
Cretaceous period 35, 61, 64–68
critical density 28
crocodiles 63
crop growing
  energy requirements 164
  perennial forms 169
Crossopterygii 59–60
crustaceans 178, 183
Crutzen, Paul 193
cyanobacteria 39–40, 44, 48–49, 54, 178
cynodont 60
cytoplasm 43

Daisyworld 208–09
dark energy 27
dark matter 25–27
Darlington, Cyril Dean 240
Dart, Raymond 71
Darwin, Charles 76, 223
Darwinian natural selection 76–80, 208–09, 220
Davies, Paul 5, 89, 97, 110, 216, 217–18
Dawkins, Richard 58, 208, 214, 235–37
De Broglie, Louis 105
deep ecology 156–57, 221–22
deforestation 154, 168, 170, 175, 176–77
Delsemme, Armand 6, 24, 36
dendrites 126–28
deoxyribonucleic acid (DNA) 41–42, 43, 54–55, 234, 237
  bacterial 185–86
  base sequences 73–74, 122
  human and chimpanzee 70, 73–74
  mutation in 76
Descartes, René 130, 137, 140
design 217–20
determinism, see quantum mechanics
deuterium 9, 25, 166, 215
deuterons 9, 215
Devall, Bill 156
Devonian period 35, 59–60
Dickens, Charles 152
dimethyl sulphide 203, 205
dinosaurs 62–68
diproton 215
dodo 172
double-slit experiments 93–7, 106
Driesch, Hans 14
dryland salinity 170–71
dualism 130–33, 136–37, 140–41
Dubois, Eugene 72
Dubos, René 161
dust, interstellar 20–22
Dyson, Freeman 110, 216, 217

Earth 22–24, 29, 34–38, 112
  as life support system 145
  as living organism 205–06
  atmosphere composition 200–05
Eccles, Sir John 128, 130, 133, 136–37, 139–41
echidna 64
ecology 144–45, 149, 157, 229
ecosystems 1, 157, 159, 162, 173–76, 210–11
  alien species in 175–76
  aquatic 178
  conservation 179, 222
  management 230–32
  Mediterranean 168
Ediacaran period 35, 56, 58, 60
effluent 178, 183
egg-laying
  reptiles 60, 62
  mammals 60, 64
Ehrlich, Paul 152
Eigen, Manfred 77
Einstein, Albert 100, 105, 110, 111, 227
Einstein-Podolsky-Rosen experiment 100
Eldredge, Niles 77
electric constant 112
electromagnetic radiation 3–8, 82–3, 86
electrons 7, 9–10, 13–14, 29–32, 85, 86, 91, 109
  and wave-particle duality 93–7
  elementary charge 112, 215

Emerson, Ralph Waldo 213
energy 7, 8, 12–13
  liberation in stars 15–19
  conservation 97
  forests 165
entangled states 100, 104–5, 106
environmental concern 144–47
  aesthetic dimension 146, 151
  and science 161–97
  in colonial administrations 148, 168
  in Europe 148–5
  in North America 151–521
  origins 148–53
  urban landscapes 146–47
  wildlife and wilderness 145–46, 151
environmental problems 159–60
  cultural origins 153–54
  and lifestyle 160
environmentalism 143–60
  and animal liberationism 231–32
  and theism/atheism 221–26
  attitudes to science 157–60
  present-day philosophies of 154–57
Eocene epoch 35, 61, 69
epiphenomenalism 131–32, 136, 140
estuaries 178, 204, 210
ethanol 165
Ethiopia 70–71
eubacteria 44, 54–55
eukaryotes 42, 45, 58, 209, 226
  first appearance 53–56, 60
Europe 149, 151, 219
evangelical Christians 224
evolution 33–80, 149
  evidence for 75–76, 226–27
  convergence in 79
  of consciousness 141–4
  mechanisms 75–80

Famintzin, Andreas 55
Faraday, Michael 82
farming 229–30
fatty acid 37
feminism 232–34
feminist science 233–34
ferns 62, 69, 227
ferrous iron 50
fertility (human) 189
fertilizers 150
fine-structure constant 112, 215
fish 178, 183
  in human ancestry 59–60
fisheries 173, 176, 178, 181–83
  sustainable yield 181
Flannery, Tim 154
flowering plants 69
fluoroquinoline resistance 187
Flynn, James 239
Fodor, Jerry 115
forests 165, 168, 170, 172, 173–5, 177, 179,-80 183–84, 190, 205
formaldehyde 20, 36
Fortey, Richard 57
fossil fuel 159, 162–64
fossils 38–41, 56–57, 75, 226–27

Gaia 155, 199–211
  proposed control mechanisms 202–0
  criticisms of theory 206–09
  implications for environmental concern 209–11
galaxies 1, 10–11, 19, 20, 26, 112, 215
Galileo 82
gamma-ray photons 6–8, 16–18, 30, 86
gas clouds 11–14, 20–22, 36
Gauss, Karl Friedrich 82
Gell-Mann, Murray 29
genes 235–41
  and culture 235, 240–41
  and human behaviour 238–41
  and intelligence 239–40
  and temperament 240
genetic determinism 234–41
  and human history 240
genetic engineering 166, 169, 183, 193–96
  techniques used 194
  and vitamin A 195
  for biofuel 196
genome
  human 73–74
  mouse 74
geological record 75, 227

geological time scale 35
glaciation 52–53, 57
global warming 162–67
consequences 164, 177–78
gluons 7, 30
glycine 37
God 1, 103, 149, 158, 21, 214, 218, 223–24, 227
Gödel, Kurt 123
Gödel's theorem 123–25
Golding, William 206
Goldsmith, Edward 158
Goodwin, Brian 77–78
gorilla 70
Gould, Stephen Jay 58, 77, 79, 235
graphite 20
gravitation 11–12, 20, 26–27, 110, 216
constant 112
gravitational collapse 12, 16–18
graviton 110
Gray, Tom 71
greenhouse effect 51–52, 159, 197, 201–03, 207
Gribbin, John 6
Grove, Richard 148
Guth, Alan 6

Hadean eon 35
Haeckel, Ernst 148–49, 156
Haile-Selassie, Yohannes 70
half-lives 87
Hall, Allan 38
halons 192–93, 197
Hameroff, Stuart 134–36
Hay, Peter 155
Hegel, Georg Wilhelm Friedrich 117
Heisenberg uncertainty principle 5, 91–3, 95–8, 133, 136
Heisenberg, Werner 92, 136
helium 9–11, 13–20, 25
nuclei 86, 215, 216
Henderson, Lawrence 217
herbicides 168–69, 195
herbivores 230–31
heterotrophic 44–45
hidden variable theories 90–1, 99–100, 105–7
histones 54
Holocene epoch 35
homeothermy 63, 68
hominids 70–74
Homo erectus 61, 71–72
Homo habilis 61, 71–72
Homo heidelbergensis 72
Homo neanderthalensis 73
Homo sapiens 33, 58, 61, 70, 72–75, 77, 138, 140–42, 171, 210, 241
Hox genes 57
Hoyle, Fred 216–17
Hubble, Edwin 2
Hughes, Miranda 233
Humanism 224–25
Hume, David 223
hunting 229
Hutt, Corinne 233
Hutton, James 206
hydrocarbons 20
hydrogen 9–11, 13–20, 25, 29, 36, 38, 45–47, 50, 166
fusion 13–17
ionization 14
hydrogen cyanide 20, 36
hydrogen sulphide 38, 45, 47, 50
hydrothermal vents 38, 46

ice 162, 217
icthyostegalians 59, 62
identical twin studies 238–39
industrial melanism 76
infinities 27, 218
inflation hypothesis 5–6
inflaton field 6
information processing
algorithmic/non-algorithmic 121–25, 134, 136, 138, 140
conscious and non-conscious 119–21, 126, 131
insecticides 159
insectivores 64, 68, 75
insects 62, 164
intelligence quotient (IQ) 239–41
interference patterns 93–7
invertebrates 59–60, 75, 178
iodine 205

ionization 14
iridium 65, 68
iron 15–20

Jeans, Sir James 217
Johanson, Donald 70
John Paul II, Pope 224
Johnson, Dr Samuel 11
Judaeo-Christian theism 153–54, 223, 225
Jupiter 23
Jurassic period 3, 61, 63–64

kangaroos 231
Kant, Immanuel 117
Kauffman, Stuart 77–78
Kirschvink, Joe 53
Knoll, Andrew 53, 58
Knox, Ronald 103
krill (*Euphausia superba*) 157
K-T boundary 64–68
Kyoto Agreement 197

Landau, Lev 1
Laplace, Pierre 83, 107, 109
lasers 85
Lawrence, D.H. 151
Le Chatelier, Henry 206
Leakey, Louis 71
Leakey, Mary 71
Leakey, Richard 72
Leopold, Aldo 152
Lewontin, Ricard 235
life-force 149–50
light, speed of 32, 84, 111–12, 215
lightning 36, 41
lipid bilayer 37, 42–44, 78
Liss, Peter 205
lithium 9, 25
lizards 62–63
lobe-finned fishes 59–60, 75
Locke, John 82
Lorenz, Edward 108
Lorenz, Konrad 235
Lovelock, James 199–211
Lucas, John 124
Luminet, Jean-Pierre 28
lycopods 62, 69

magnesium 15–20
male-female differences 232–33
mammals 60–61, 63–68, 228
  and consciousness 141
  egg-laying 60
  marsupial 79
  placental 61, 79
  visual pigments 69
Margenau, Henry 133
Margulis, Lynn 55, 201, 206, 208, 210
marine reserves 182
Mars 22, 200
Marsh, George Perkins 151
marsupials 175
Maruyama, Shigenori 39
mass energy 111
mass extinctions 35
materialism 34, 131, 214–21
mathematics 215, 218–19
Mauna Loa 162–63
Maxwell, James Clerk 82
May, Sir Robert 171
Mayr, Ernst 77
McDonagh, Fr. Sean 224
Mediterranean region 154, 168
melanoma 192
memes 214
Mereschkowski, Konstantin 55
mesons 7, 9, 30
Mesoproterozoic era 35
Mesozoic era 35, 64
metal sulphides 38
metamorphic rock 38–39
metaphysics 34, 80, 214, 218, 222–23
methane 20, 36, 38, 39, 40, 45, 50
  consumption by bacteria 52
  from wetlands 204–05
  in atmosphere 200, 202
methanol 36
microtubules 134–35, 138
microwaves 3, 10, 12, 26–28
Milky Way 19
Miller, Stanley 36
minimum tillage 168
Miocene epoch 35, 61

mitochondria 43, 53
  origin 55–56
moa 172
Mojzsis, Stephen 39
molecular fossils 38, 53
Molina, Mario 193
molluscs 58
monism 149
monkeys 61, 69
monotheism 219
Moon 22–24
morphogenesis 78
Morris, Desmond 235
Morris, Simon Conway 79, 222
Morris, William 151
Muir, John 15, 156
multicellular organisms, origin 56–57, 60, 226
multiverse 218
musical pitch perception 238
mutation 76, 185

Naess, Arne 156–57
Napier, John 71
NASA 12, 24, 28, 200
national parks 151
nature 225–26, 229
  beauty of 147, 151
  poetry 151
nature versus nurture 235–41
Neanderthal humans 72–73
neocortex 141
neon 15–19
Neoproterozoic era 35
Neptune 23
nerve fibres 126–30
neuronal correlates of consciousness 131
neurons 116, 126–30, 135, 14, 220
neurophysiology 125–30, 131
neurotransmitters 129–30
neutrinos 7–10, 31–32
neutrons 7, 9, 25, 29–32
New Guinea 180, 190
Newton, Sir Isaac 82, 109, 110, 219
Newtonian mechanics 90
nitrate 36, 203
nitrogen 20, 24, 31, 36
  in water bodies 178
North America 151, 156, 173, 239
nuclear fusion 13–17, 85, 166
nuclear power 159, 166–67
nucleic acids 36–37, 41–42
nucleons 7–9, 29–32, 215
nucleosynthesis 14–19, 25, 216
nucleotides 37, 41
nucleus 29–30, 87–9, 95
nucleus 54

ocean 36, 41, 46, 78, 164, 183
  and global carbon cycle 202–04, 20
  temperature 178–79
Oligocene epoch 35, 61, 69
Opuntia cactus 180
orange roughy (*Hoplostethus atlanticus*) 181–82
orang-utan 70
Ordovician period 35, 59–60
organic farming 158
overgrazing 154, 168, 175
overpopulation 152
oxygen 15–20, 59, 216–17
  and ozone formation 51–2, 190–91
  liberation in photosynthesis 48–52, 204
  atmospheric concentration 51, 57, 200, 204–05
  and multicellularity 57
  depletion in polluted waters 178
ozone layer 51–52, 159–60, 190–93
  hole in 191, 193, 197

Pacific region 154, 168, 171–72, 183
palaeosol 52
Paleocene epoch 35, 61, 68
Paleoproterozoic era 35
Paleozoic era 35
Pan species 773
panpsychism 132–33, 136, 140
pantheism 149
passenger pigeon 172–73
penicillin 184
Penrose, Roger 124, 134–38
perennial plants 165, 169

Periodic Table 17, 19, 25
Perlmutter, Saul 26
Permian period 35, 60, 62–63
Pesticides 152
Phanerozoic eon 35
Philosophy 214
  Greek 82
  realist 82, 116–1
  idealist 116–18
phosphate 36, 203
  in water bodies 178
photoelectricity 93
photons 3–4, 6–8, 30, 85, 89, 100
  and wave-particle duality 93–7
photosynthesis 43, 46–50, 68, 202–04
phylum 58, 62
physical constants 2, 111–12, 209, 215, 218
physical laws 2, 214–15, 218–19
physics, classical 82–5
phytoplankton 203–05, 208
pigment 46–48
Pilbara region 39
pilot wave 105–7
Planck, Max 83–5
Planck's constant 84, 98, 112–14
planet formation 20–3
planetary atmospheres 200
planetary embryos 22
planetesimals 22–24, 36
plankton 211
plant breeding 237–38
plasma 10–11
platypus 64
Pleistocene epoch 35, 61
Pliocene epoch 35, 61
Poincaré, Henri 108
pollution 145, 152, 159–60, 177–78, 183, 189–90
polysaccharides 41, 165
Popper, Sir Karl 130, 136–37
population growth 188–89, 225
positrons 7, 13, 31–32, 86
predators and prey 229, 231
primates 61, 68–69
prokaryotes 43–49, 54, 58, 79, 209, 226
prosimians 61, 69, 75
proteins 36, 41–42
Proterozoic eon 35
protons 7, 9, 13–14, 25, 29–32, 85, 215
mass ratio 112
protozoa 53
purines 36–37
pyrimidines 36–37

quantum decoherence 102, 1, 34–35, 137
quantum electrodynamics 30
quantum jumps 136
quantum mechanics 30, 83–107, 117
and consciousness 133–38, 220
and determinism 86–91, 104
and gravitation 110–11, 134
and reality 103–5
non-locality 98–100
spin 98–100
measurement problem 100–3, 105
quantum tunnelling 88–9
quantum vacuum 5
quantum-coherent state 135–36
quarks 7, 29–30
Quaternary period 35

rabbit 175
  virus control 180
racism 227, 235–37
radio astronomy 20
radioactive decay 86–8, 104
radium 87
recombination 10–11
red fox 175, 231
Redman, Stephen 133
relativity theory 26, 83, 110
religion 214, 219, 223–25, 235
reptiles 60, 62–64
  diapsids and synapsids 6
  mammal-like 63, 75
  reptilian brain 141
resource depletion 145, 155, 171–84, 225
ribonucleic acid (RNA) 41–42
rivers 178, 190
rodents 68
Rose, Stephen 235

Rowland, F. Sherwood 193
rubber vine (*Cryptostegia grandiflora*) 176, 181
Ruskin, John 151
Russell, Bertrand 213
Russell, Michael 38

Sahelanthropus tchadensis 70
salinization 159, 169–71
  in early civilizations 170
Salvinia molesta 180
saprophytes 49
Saturn 23
Schauder, Rolf 40
Schmidt, Brian 26
Schneegurt, Mark 40
Schopenhauer, Arthur 117
Schrödinger wave equation 87–9, 95, 101–2, 106, 112–14, 136
Schrödinger, Erwin 105
Schuster, Peter 77
Schwartz, Stephen 208
sedimentary rocks 38–9
Segerstrale,Ullica 235
Sessions, George 156
sewage 159, 178
sexism 227
Shakespeare, William 81, 115
Shark Bay (WA) 49
Shaw, George Bernard 149
silicates 20, 22–23, 36
  weathering in soils 207–08, 211
silicon 15–20
Silurian period 35, 59–60
Singer, Peter 229–30
Singh, Simon 6
Snowball glaciation 53, 57
social ecology movement 155–56
sociobiology 235–36
sociology 235
soil erosion 154, 167–69, 175, 177, 183
solar energy 164–65
solar radiation 84
Solar System 20–24
sound waves 11
South America 177
Southern Ocean 157, 178, 181
space-time dimensions 111
speciation 77
species 156–57
  extinction 173
  value 222–23
speciesism 227
Squires, Euan 103, 106, 124
Staphylococcus aureus 184
Stapp, Henry 136–37
stars 1, 112
  formation 12–13, 21
  nuclear reactions in 13–17
state vector reduction 101–2
  and consciousness 134–38
Steiner, Rudolf 150
stellar evolution 15–19
steranes 53
stratosphere 190
streptomycin 184, 187
string theory 110–11, 218
stromatolites 39, 44, 46, 48–49
strong nuclear force 30, 88, 215–16
sugars 36, 46
sulphate 36, 205
sulphur 24
  transfer from sea to land 205
sulphur dioxide 20, 177, 208
Sun 13, 21–23, 29, 34, 84–5, 216
  luminosity 51, 201, 203, 207
superconductivity 85, 135
supernovae 16, 18–19, 20, 26, 34
superposition, principle of 95, 102, 104, 134, 136–37
symbiosis 54–56
synapse 126–30, 135, 140
  presynaptic vesicular grid 128–29, 133, 136
synaptic knob 126–30, 133, 136
synaptic vesicles 128–29, 133, 136, 220
Tasmanian tiger (*Thylacinus cynocephalus*) 79, 172
taxol 174–75
tectonic movement 23
Tegmark, Max 135
Tertiary period 35, 64–70

tetrapods 59, 75, 78
theism 213–26
  and environmentalism 221–26
Thoreau, Henry David 151
thorium 25
Tipler, Frank 5, 215
Tobias, Philip 71
transistors 85
trees 175
  as biofuel source 165
Triassic period 35, 60, 62–64
tropical rainforests 148, 168, 174–75, 177, 179, 204, 210
  and hydrological cycle 177
troposphere 190
tuberculosis 184
tubulin 135, 220
Tudge, Colin 226
tuna 176
turtles 62–63
twins 238–39

Ueno, Yuichiro 39
Ultraviolet (UV) 14, 30, 38
Universe 1–32
  chemical composition 25
  design 217–20
  expansion 2–4, 8–10, 26–27
  mass density 26–27
  origin 5–6
  size 27–28
uracil 37
uranium 25, 86–7, 90–1, 166
Uranus 23
Urey, Harold 36

vacuum 5
vacuum energy 27
vancomycin 184
vegetarianism 227, 230
Venus 200, 201
Vernadsky, Vladimir Ivanovich 206
virtual particles 5, 98
visual pigments 64, 69
vitalism 149
volcanoes 53, 64, 201, 203
von Neumann, John 102

Wächterhäuser, Günter 38
Walker, Alan 72
Walker, James 207
Walter, Malcolm 39
water 20, 24
  anomalous properties of 217
  as hydrogen donor 47–48
Watson, Andrew 206, 208
wave energy 87
wave packet 89, 94
wavefunction 87–9, 95–6, 99, 104–5
  and consciousness 134–37
  collapse 101–3, 105, 134
wave-particle duality 93–7, 104
  of fullerene molecules 94, 106
weather forecasts 108–9
Weinberg, Stephen 6
Weisskopf, Victor 6
wetlands 183, 204, 210
White, Tim 70
Whitehead, Alfred North 132
Wigner, Eugene 102, 215
wilderness conservation 145–46
wildlife 229
  conservation 145–46, 210
Wilkinson Microwave Anisotropy Probe 11–12, 26–28
Wilson, Edward 171, 235–36, 241
Wordsworth, William 151
Wulst 141
Wynne, Clive 139

X-rays 30

Zeilinger, Anton 94
Zweig, George 29