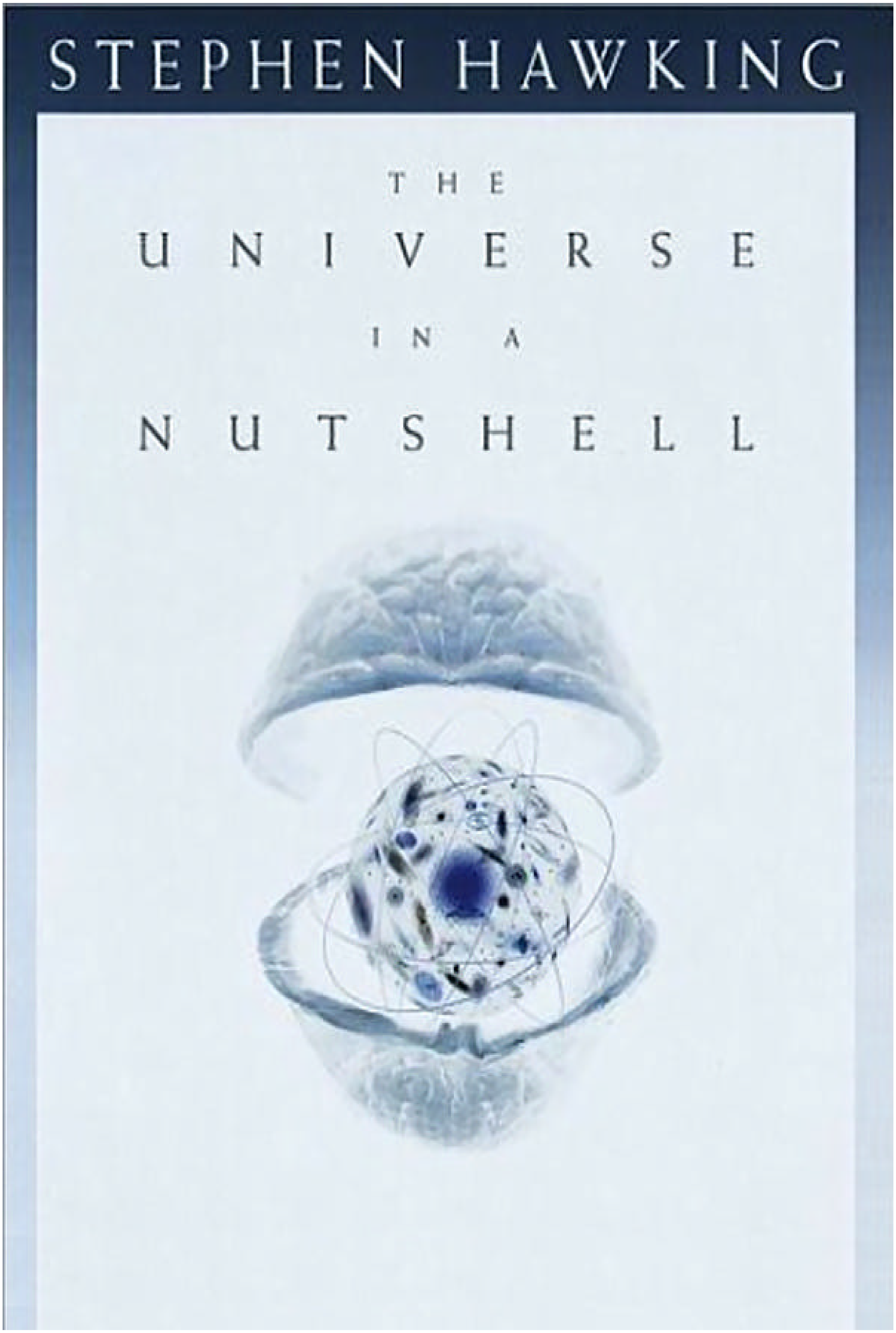
STEPHEN HAWKING
THE
UNIVERSE
IN A
NUTSHELL

The Universe in a Nutshell

by Stephen Hawking

transcrisa dupa un format audio .mp3 de Mihai Rusie

- **Hardcover:** 224 pages
- **Publisher:** Bantam; 1st edition (November 6, 2001)
- **Language:** English

ISBN: 0-553-80202-X

Introduction

I hadn't expected my popular book "A Brief History of Time" to be such a success; it was on the London Sunday Time's bestseller list for over 4 years, which is longer than any other book has been, and remarkable for a book on science, it was not easy going.

After that, people kept asking when I would write a sequel; I resisted because I didn't want to write a "son of", or "a slightly longer brief history of time", and because I was busy with research.

But I've come to realize that there is room for a different kind of book that might be easier to understand.

"A brief history of time" was organized in a linear fashion, with most chapters following and logically depending on the preceding chapters; this appealed to some readers, but others got stuck in the early chapters, and never reached the more exciting material later on; by contrast, the present book is more like a tree; chapters 1 and 2 form a central trunk from which the other chapters branch off. The branches are fairly independent with each other, and can be tackled in any order after the central trunk. They correspond to areas I've worked on or thought about since the publication of "A brief history of time", thus, they present a picture of some of the most active fields of current research. Within each chapter I've also tried to avoid a single linear structure.

In 1988, when " A brief history of time " was first published, the ultimate "theory of everything" seemed to be just over the horizon. How is the situation changed since then? Are we any closer to our goal? As will be described in this book, we have advanced a long way since then; but it's a long going journey still, and the end is not yet in sight. According to the old saying: "it's better to travel hopefully, than to arrive".

Our quest for discovery fuels our creativity in all fields, not just science. If we reached the end of the line, the human spirit will shrivel and die. But I don't think we'll ever stand still, we should increase in complexity, if not in depth, and should always be the centre of an expanding horizon of possibilities. I want to share my excitement of the discoveries that are being made and the picture of reality that is emerged. I've concentrated on areas I've worked on myself, for a greater feeling of immediacy. The details of the work are very technical, but I believe the broad ideas can be conveyed without a lot of mathematical baggage. I just hope I've succeeded.

I had a lot of help with this book; I would mention in particular, Thomas Herthog (?) and Neil Schearer (?) for assistance with the figures, captions and boxes, and Harrison Kitty-Fergusson (?) who edited the manuscript, or more accurately, the computer files, because all I write is electronic; and Phillip Dunn (?) of the book laboratory and Moon-Runner Design, who created the illustrations.

But beyond that, I want to thank all those who have made it possible for me to lead a fairly normal life, and carry on scientific research.

Without them this book could not have been written.

Stephen Hawking, Cambridge, May the 2nd, 2001.

Chapter 1

A Brief History of Relativity

1.1 *How Einstein laid the foundations of the two fundamental theories of the 20-th century: General Relativity, and Quantum Theory*

Albert Einstein, the discoverer of the special and general theories of relativity, was born in (...) in Germany, in 1879; but the following year, the family moved in Munich, where his father, Hermann, and uncle Jacob, set up a small and not very successful electrical business.

Albert was no child prodigy, but claims that he did poorly at school seem to be an exaggeration.

In 1894 his father's business failed, and the family moved to Milan. His parents decided he should stay behind to finish school, but he didn't like its authoritarianism, and within months he left to join his family in Italy. He led to complete his education in Zurich, graduating from the prestigious Federal Polytechnical School, known as the ETH, in 1900.

His argumentative nature and dislike of authority did not endear him to the professors of the ETH, and none of them offered him the position of assistant, which was the normal route to an academic career. Two years later he finally managed to get a junior post at the Swiss Patent Office in Bern. It was while he held this job that in 1905 he wrote three papers that both established him as one of the word's leading scientists and started two conceptual revolutions; revolutions that changed our understanding of time, space and reality itself.

Towards the 19-th century scientists believed they were close to a complete description of the universe. They imagined that space was filled by a continuous medium called the Ether. Light rays and radio signals were waves in this ether just as sound in pressure waves in air. All that was needed for a complete theory were careful measurements of the elastic properties of the ether.

In fact, anticipating such measurements the Jefferson Lab at Harvard University was built entirely without iron nails, so as not to interfere with delicate magnetic measurements. However, the plan has forgot that the reddish-brown bricks, of which the lab and most of Harvard are built, contain large amounts of iron. The building is still in use today, although Harvard is still not sure how much weight the library floor without the iron nails will support.

By the century's end discrepancies in the idea of an all-pervading ether began to appear. It was expected that light would travel at a fixed speed through the ether, but that if you were travelling through the ether in the same direction of the light, its speed would appear lower, and if you were travelling in the opposite direction of the light, its speed would appear higher. Yet, a series of experiments failed to support this idea; the most careful and accurate of these experiments was carried out by Albert Michelson and Edward Morely at the Case School of Applied Science in Cleveland - Ohio, in 1887. They compared the speed of light in two beams at right angles to each other.

As the Earth rotates on its axis and orbits the Sun, the apparatus moves through the ether with varying speed and direction; but Michelson and Morely found no daily or yearly differences between the two beams of light. There was as if light always travelled at the same speed relative to where one was, no matter haw fast and in which direction one was moving. Based on the Michelson-Morely experiment the Irish physicist George Fitzgerald , and the Dutch physicist Hendrich Lorentz, suggested that bodies moving through the ether would contract, and that clocks would slow down.

This contraction and slowing down of clocks would be such that people would all measure the same speed for light, no matter how they were moving with respect to the ether. Fitzgerald and Lorentz still regarded ether as a real substance. However, in a paper written in June 1905, Einstein pointed out that if one could not detect whether or not one was moving through the space, the notion of an ether was redundant; instead, he started from the postulate that the laws of science should appear the same to all freely moving observers. In particular, they should all measure the same speed for light, no matter how fast they were moving.

The speed of light is independent to their motion and is the same in all directions. This required abandoning the idea that there is an universal quantity called time that all clocks would measure. Instead, everyone would have his or her own personal time. The times of two people would agree if the people were at rest with respect to each other, but not if they were moving.

This is being confirmed by a number of experiments including one in which two accurate clocks were flown in opposite directions around the world and returned showing very slightly different times. This might suggest that if one wanted to live longer, one should keep flying to the East, so that the plane's speed is added to the Earth rotation. However, the tiny fraction of a second one would gain, would be more than cancelled by eating airline meals.

1.2 *The twins paradox*

In the theory of relativity, each observer has his own measure of time; this could lead to the so-called twins paradox.

One of the pair of twins leaves on a space journey during which he travels close to the speed of light, while his brother remains on Earth.

Because of his motion, time runs more slowly in the spacecraft, as seen by the Earth-bound twin; so on his return, the space traveller will find that his brother has aged more than himself. Although it seems against common sense, a number of experiments have implied that in this scenario, the travelling twin would indeed be younger.

A spaceship passes Earth from left to right at four fifths the speed of light; a pulse of light is emitted at one end of the cabin and reflected at the other end; the light is observed by peopled on Earth and on the spaceship. Because of the motion of the spaceship, they would disagree about the distance the light has travelled in reflecting back. They must therefore also dis agree about the time the light has taken. According to Einstein's postulate that the speed of light is the same for all freely moving observers.

Einstein's postulate, that the laws of nature should appear the same to all freely moving observers was the foundation of the theory of relativity, so called because it implied that only relative motion was important. Its beauty and simplicity convinced many thinkers, but there remained a lot of opposition. Einstein had overthrown two of the absolutes of 19-th century's science: absolute rest - as represented by the ether, and absolute, or universal time - which all clocks would measure. Many people found this an unsettling concept: did it implied to us that everything was relative, that there were no absolute moral standards? This unease continued throughout the 1920's and 1930's.

When Einstein was awarded the Nobel Prize, in 1921, the citation was "for important", but, by his standard, "comparatively minor work" also carried out in 1905; it made no mention of relativity, which was considered too controversial. I too get 2 or 3 letters a week telling me: "Einstein was wrong (...)". Nevertheless, the theory of relativity is now completely accepted by the scientific community, and its predictions have been verified in countless applications.

A very important consequence of relativity is the relation between mass and energy; Einstein's postulate that the speed of light should appear the same for everyone implied that nothing can be moving faster than light. What happens is, that as one uses energy to accelerate anything, whether a particle or a spaceship, its mass increases, making it harder to accelerate it further.

To accelerate a particle, at the speed of light would be impossible, because it would take an infinite amount of energy; mass and energy are equivalent, as its summed up in Einstein's famous equation "$E = mc^2$". This is probably the only equation in physics to have recognition on the street. Among its consequences was the realisation that if the nucleus of an Uranium atom fissions into two nuclei, with slightly less total mass, this would release a tremendous amount of energy.

In 1939, as the prospect of another WW (...) a group of scientists who realized these implications, persuaded Einstein to overcome his pacifist scruples, and add his authority to a letter to pres. Roosevelt, urging the USA to start a program of nuclear research. This led to the Manhattan project, and ultimately to the bombs that exploded over Hiroshima and Nagasaki, in 1945. Some people blamed the atom bomb on Einstein, because he discovered the relationship between mass and energy, but that's like blaming Newton for causing airplanes to crash, because he discovered gravity. Einstein himself took no part in the Manhattan project, and was horrified by the dropping of the bombs.

1.3 *Einstein's prophetic letter to pres. Roosevelt in 1939*

In the course of the last 4 months it is been made probable through the work of Joliot (?) in France, as well as Fermi and Zillag in America, that it may become possible to set up a nuclear chain reaction in a large mass of Uranium by which vast amounts of power, and large quantities of new Radium-like elements would be generated. Now it appears almost certain that this could be achieved in the immediate future.

This new phenomenon would also lead to the construction of bombs and it is conceivable, though much less certain, that extremely powerful bombs of a new type may thus be constructed.

After his groundbreaking papers in 1905, Einstein's scientific reputation was established; but it was not until 1909 that he was offered a position at the University of Zurich that enabled him to leave the Swiss Patent Office. Two years later he moved to the German University in Prague, but he came back to Zurich in 1912, this time to the ETH, despite the anti-Semitism that was common in much of Europe, even in the universities, he was now an academic hot property. Offers came in from Vienna and Utrecht, but he chose to accept a research position at the Prussian Academy of Sciences in Berlin, because it freed him from teaching duties. He moved to Berlin in April 1914 and was joined shortly after by his wife and two sons. The marriage had been in a bad way for some time, however, and his family returned to Zurich; although he visited them occasionally, he and his wife were eventually divorced. Einstein later married his cousine Elsa, who lived in Berlin.

The fact that he spent the war years as a bachelor, without domestic commitments, may be one reason why this period was so productive for him, scientifically.

Although the theory of relativity fitted well with the laws that govern electricity and magnetism, it was not compatible with Newton’s law of gravity. This law said that if one changed the distribution of matter in one region of space, the change in the gravitational field would be felt instantaneously everywhere else in the universe. Not only this would mean one could send signals faster than light, something that was forbidden by relativity, in order to know what instantaneous meant, it also required the existence of absolute or universal time, which relativity had abolished in favour of personal time. Einstein was aware of this difficulty in 1907, while he was still at the Patent Office in Bern, but it was not until he was in Prague, in 1911, that he began to think seriously about the problem. He realized that there's a close relationship between acceleration and the gravitational field. Someone inside a closed box, such as an elevator, could not tell whether the box was at rest in the Earth's gravitational field, or is being accelerated by a rocket in free space.

Of course that this was before the age of "Star Trek", and so Einstein thought of people in elevators, rather than spaceships. But one cannot accelerate or fall freely very far in an elevator, before disaster strikes. If the Earth were flat, one could equally well say that the apple fell on Newton's head because of gravity, or because Newton, and the surface of the Earth, were accelerating upward. This equivalence between acceleration and gravity didn't seem to work for around Earth, however; people on the opposite sides of the world would have to be accelerating in opposite directions but staying at a constant distance from each other. But on his return to Zurich, in 1912, Einstein had the brainwave of realizing that the equivalence would work if the geometry of space-time was curved and not flat, as had been assumed had it to. His idea was that mass and energy would warp space-time in some manner yet to be determined. Objects such as apples and planets would try to move in straight line through space-time, but their paths would appear to be bent by gravitational field, because space-time is curved.

With the help of his friend, Marcell Grassman, Einstein studied the theory of curved spaces and surfaces, that have been developed earlier by George Friedrich Riemann, as a piece of abstract mathematics. Riemann had had no thought that it would be relevant to the real world. Einstein and Grassman wrote a join paper in 1913, in which they put forward the idea that what we think of as gravitational forces are just an expression of the fact that space-time is curved. However, because of a mistake by Einstein, who was quite human and fallible, they weren't able to find the equations that relate the curvature of space-time to the mass and energy in it.

Einstein continued to work on the problem in Berlin, undisturbed by domestic matters, and largely unaffected by the war, until he finally found the right equations in November 1915. He had discussed his ideas with the mathematician David Hilbert, during the visit to the University of (Gertingham...?) in the summer of 1915. And Hilbert, independently found the same equations a few days before Einstein; nevertheless, as Hilbert himself admitted, the credit for the new theory belonged to Einstein. It was his idea to relate gravity to the warping of space-time. It's a tribute to the civilized state of Germany this period, that such scientific discussions and exchanges could go on undisturbed, even on war time. It was a sharp contrast to the nazzies 20 years later.

The new theory of space-time was called "general relativity", to distinguish it from the original theory without gravity, which was now known as the "special relativity". It was confirmed in a spectacular fashion in 1919, when a British expedition to West Africa, observed a slight bending of light passing near the Sun, during an eclipse. Here was direct evidence that space and time are warped, and it spurred the greatest change in our perception of the Universe in which we live, since Euclid wrote his "Elements of Geometry" around 300 b.C.

Einstein's theory of relativity transforms space and time from a passive background in which events take place to active participants in the dynamics of the universe.

This led to a great problem that remains at the forefront of physics in the 21-st century. The Universe is full of matter, and matter warps space-time in such a way that bodies fall together. Einstein found that his equations didn't have a solution that described the universe unchanging in time. Rather than giving up such an everlasting universe, which he and most other people believed in, he funched the equations by adding a term called "the cosmological constant", which warps space-time in the opposite sense, so that bodies move apart. The repulsive effect of the cosmological constant could balance the attractive effect of the matter, thus allowing a static solution for the universe. This was one of the missed opportunities of theoretical physics. If Einstein had stuck with his original equations, he could have predicted that the universe would be either expanding or contracting; as it was, the possibility of a time dependent universe wasn't taken seriously, until observations in the 1920's by the 100 inch telescope on Mount Wilson. These observations revealed that the farther other galaxies are from us, the faster they're moving away; the universe is expanding, with the distance between any two galaxies steadily increasing with time.

This discovery removed the need for a cosmological constant in order to have a static solution for the universe. Einstein later called the cosmological constant "the greatest mistake of his life". However, it now seems that it may not have been a mistake after all; recent observations, described in chapter 3, suggest that it may indeed be a small there is a cosmological constant.

General relativity completely changed the discussion of the origin and fate of the universe. A static universe could have existed forever, or could have been created in its present form at some time in the past. However, if galaxies are moving apart now, it means that they must have been closer together in the past. About $15*10^9$ years ago, they would all have been on top of each other, and the density would have been very large. This stage was called the "primeval atom", by the catholic priest Georges Lemaître, who was the first to investigate the origin of the universe, that we now call the "big bang".

Einstein seems never to have taken the big bang seriously; he apparently thought that a simple model of a uniformly expanding universe would break down if one followed the motions of the galaxies back in time; and that the small sideways velocities of the galaxies would cause them to miss each other. He thought that the universe might have had a previous contracting phase, with a bounce into the present expansion, at a fairly moderate density. However, we now know, that in order for nuclear reactions in the early universe to produce the amounts of light elements we now observe around us, the density must have been at least 10 tons per cubic inch (@ $4*10^6 * H_2O$), and the temperature $10*10^9$ degrees. Further observations of the microwave background indicate that the density was probably once a trillion trillion trillion trillion trillion trillion (1 with 72 zeros after it) tons per cubic inch. We also now know that Einstein's general theory of relativity does not allow the universe to bounce from a contracting phase to the present expansion.

As will be discussed in chapter 2, Roger Penrose and I, were able to show that general relativity predicts that the universe began in the big bang. So Einstein's theory does imply that time has a beginning, although he was never happy with the idea. Einstein was even more reluctant to admit that general relativity predicted that time would come to an end for massive stars, when they reach the end of their life, and no longer generated enough heat to balance the force of their own gravity, which was trying to make them smaller. Einstein thought that such stars would settle down to some final state; but we now know that there are no final state configurations for stars of more than twice the mass of the Sun. Such stars would continue to shrink until they become "black holes", regions of space-time that are so warped that light cannot escape form them.

When a massive star exhausts its nuclear fuel, it will lose heat and contract; the warping of space-time would become so great that a black hole would be created, from which light cannot escape. Inside the black hole, time would come to an end. Penrose and I showed that general relativity predicted that time would come to an end inside a black hole, both for the star, and for any unfortunate astronaut who happened to fall into it. But both the beginning and the end of time would be places where the equations of general relativity could not be defined. Thus the theory could not predict what should emerge from the big bang. Some saw this as an indication of God's freedom to start the universe of in any way God wanted; but others, including myself, felt that the beginning of the universe should be governed by the same laws that held at other times.

We've made some progress toward this goal, as will be described in chapter 3, but we don't yet have a complete understanding of the origin of the universe.

The reason general relativity broke down at the big bang was that it was not compatible with quantum theory, the other great conceptual revolution of the early 20-th century. The first step towards quantum theory had come in 1900 when Max Plank, in Berlin, discovered that the radiation from a body that was glowing red-hot was explainable if light could be emitted or absorbed, only if it came in discrete packets, called quanta. In one of his ground-breaking papers written in 1905, when he was at the Patent Office, Einstein showed that Plank's quanta hypothesis could explain what is called the photo-electric effect - the way certain metals give off some electrons when light falls on them. This is the basis of modern light detectors in television cameras, and it was for this work that Einstein was awarded the Nobel price for physics. Einstein continued to work on a quantum idea into the 1920's but he was deeply disturbed by the work of Werner Heisenberg in Copenhagen, Paul Dirac in Cambridge, and Erwin Schroedinger in Zurich, who developed a new picture of reality, called quantum mechanics. No longer tiny particles had a definite position and speed, instead, the more accurately one determined a particle's position, the less accurately one could determine it's speed, and vice versa.

Einstein was horrified by this random, unpredictable element in the basic laws, and never fully accepted quantum mechanics. His feelings were expressed in his famous dicton "God does not play dice". Most other scientists, however, accepted the validity of the new quantum laws, because of the explanations they gave for a whole range of previously unaccounted for phenomena, and the excellent agreement with observations. They are the basis of modern developments in chemistry, molecular biology and electronics, and the foundation of the technology that has transformed the world in the last 50 years.

In December 1932, aware of the nazzis and Hitler were about to come to power, Einstein left Germany, and four months later, renounced his citizenship, appending the last 20 years of his life at the Institute for Advanced Studies in Princeton - New Jersey.

In Germany, the nazzis launched a campaign against Jewish science, and there were many German scientists who were Jews; this is part of the reason that Germany was not able to build an atomic bomb. Einstein and relativity were principal targets of this campaign. When told of the publication of a book entitled "100 authors against Einstein", he replied: "Why 100? If I were wrong, 1 would have been enough".

After the WW2, he urged the allies to set up a world government to control the atomic bomb. In 1948 he was offered the presidency of the new state of Israel, but turned it down. He once said: "Politics is for the moment. But an equation is for eternity".

The Einstein's equations of general relativity are his best epitaph in memorial. They should last as long as the universe.

The world has changed far more in the last 100 years than in any previous century; the reason is not the new political or economic doctrines, but the vast developments in technology, made possible by advances in basic sciences. Who better symbolizes those advances than Albert Einstein.

Chapter 2

The Shape of Time

2.1 ***Einstein's general relativity gives time a shape***
How this can be reconciled with quantum theory

What is time? Is it an ever-rolling stream that bears all our dreams away, as the old hymn says? Or is it a railroad track? Maybe it has loops and branches, so you can keep going forward and yet return to an earlier station on the line.

The 19-th century author Charles Lamb wrote: "Nothing puzzles me like time and space. And yet nothing troubles me less than time and space, because I never think of them."

Most of us don't worry about time and space most of the time, whatever that may be; but we all do wonder sometimes what time is, how it began, and where it is leading us.

Any sound scientific theory, whether of time or of any other concept, should in my opinion be based on the most workable philosophy of science: the positivist approach put forward by Karl Popper and others. According to this way of thinking, a scientific theory is a mathematical model that describes and codifies the observations we make. A good theory will describe a large range of phenomena on the basis of a few simple postulates and will make definite predictions that can be tested. If the predictions agree with the observations, the theory survives that test, though it can never be proved to be correct.

On the other hand, if the observations disagree with the predictions, one has to discard or modify the theory. (At least, that is what is supposed to happen. In practice, people often question the accuracy of the observations and the reliability and moral character of those making the observations.) If one takes the positivist position, as I do, one cannot say what time actually is. All one can do is describe what has been found to be a very good mathematical model for time and say what predictions it makes.

Isaac Newton gave us the first mathematical model for time and space in his PRINCIPIA MATHEMATICA, published in 1687. Newton occupied the Lucasian chair at Cambridge that I now hold, though it wasn't electrically operated in his time. In Newton's model, time and space were a background in which events took place but which weren't affected by them. Time was separate from space and was considered to be a single line, or railroad track, that was infinite in both directions. Time itself was considered eternal, in the sense that it had existed, and would exist, forever.

By contrast, most people thought the physical universe had been created more or less in its present state only a few thousand years ago. This worried philosophers such as the German thinker Immanuel Kant. If the universe had indeed been created, why had there been an infinite wait before the creation? On the other hand, if the universe had existed forever, why hadn't everything that was going to happen already happened, meaning that history was over? In particular, why hadn't the universe reached thermal equilibrium, with everything at the same temperature?

Kant called this problem an "antimony of pure reason", because it seemed to be a logical contradiction; it didn't have a resolution. But it was a contradiction only within the context of the Newtonian mathematical model, in which time was an infinite line, independent of what was happening in the universe.

However, as we saw in chapter 1, in 1915 a completely new mathematical model was put forward by Einstein: the general theory of relativity. In the years since Einstein's paper, we've have added a few ribbons and bows, but our model of time and space is still based on what Einstein proposed. This and the following chapters will describe how our ideas have developed in the years since Einstein's revolutionary paper. It has been a success story of the work of a large number of people, and I'm proud to have made a small contribution.

General relativity combines the time dimension with the three dimensions of space to form what is called space-time. The theory incorporates the effect of gravity by saying that the distribution of matter and energy in the universe warps and distorts space-time, so that it is not flat. Objects in this space-time try to move in straight lines, but because space-time is curved, their paths appear bent. They move as if affected by a gravitational field.

As a rough analogy, not to be taken too literally, imagine a sheet of rubber. One can place a large ball on the sheet to represent the Sun. The weight of the ball will depress the sheet and cause it to be curved near the Sun. If one now rolls little ball bearings on the sheet, they won't roll straight across to the other side but instead will go around the heavy weight, like planets orbiting the Sun.

The analogy is incomplete because in it only a two-dimensional section of space (the surface of the rubber sheet) is curved, and time is left undisturbed, as it is in Newtonian theory. However, in the theory of relativity, which agrees with a large number of experiments, time and space are inextricably tangled up. One cannot curve space without involving time as well. Thus time has a shape. By curving space and time, general relativity changes them from being a passive background against which events take place to being active, dynamic participants in what happens. In Newtonian theory, where time existed independently of anything else, one could ask: "What did God do before He created the universe?" As Saint Augustine said, one should not joke about this, as did a man who said, "He was preparing Hell for those who pry too deep". It is a serious question that people have pondered down the ages. According to Saint Augustine, before God made heaven and earth, He did not make anything at all. In fact, this is very close to modern ideas.

In general relativity, on the other hand, time and space do not exist independently of the universe or of each b68 other. They are defined by measurements within the universe, such as the number of vibrations of a quartz crystal in a clock or the length of a ruler. It is quite conceivable that time defined in this way, within the universe, should have a minimum or maximum value; in other words, a beginning or an end. It would make no sense to ask what happened before the beginning or after the end, because such times would not be defined.

It was clearly important to decide whether the mathematical model of general relativity predicted that the universe, and time itself, should have a beginning or end. The general prejudice among theoretical physicists, including Einstein, held that time should be infinite in both directions. Otherwise, there were awkward questions about the creation of the universe, which seemed to be outside the realm of science. Solutions of the Einstein equations were known in which time had a beginning or end, but these were all very special, with a large amount of symmetry. It was thought that in a real body, collapsing under its own gravity, pressure or sideways velocities would prevent all the matter falling together to the same point, where the density would be infinite. Similarly, if one traced the expansion of the universe back in time, one would find that the matter of the universe didn't all emerge from a point of infinite density. Such a point of infinite density was called a singularity and would be a beginning or an end of time.

In 1963, two Russian scientists, Evgenii Lifshitz and Isaac Khalatnikov, claimed to have proved that solutions of the Einstein equations with a singularity all had a special arrangement of matter and velocities. The chances that the solution representing the universe would have this special arrangement were practically zero. Almost all solutions that could represent the universe would avoid having a singularity of infinite density:

Before the era during which the universe has been expanding, there must have been a previous contracting phase during which matter fell together but missed colliding with itself, moving apart again in the present expanding phase. If this were the case, time would continue on forever, from the infinite past to the infinite future.

Not everyone was convinced by the arguments of Lifshitz and Khalatnikov. Instead, Roger Penrose and I adopted a different approach, based not on a detailed study of solutions but on the global structure of space-time. In general relativity, space-time is curved not only by massive objects in it but also by the energy in it. Energy is always positive, so it gives space-time a curvature that bends the paths of light rays toward each other.

Now consider our past light cone, that is, the paths through space-time of the light rays from distant galaxies that reach us at the present time. In a diagram with time plotted upward and space plotted sideways, this is a cone with its vertex, or point, at us.

As we go toward the past, down the cone from the vertex, we see galaxies at earlier and earlier times. Because the universe has been expanding and everything used to be much closer together, as we look back further we are looking back through regions of higher matter density. We observe a faint background of microwave radiation that propagates to us along our past light cone from a much earlier time, when the universe was much denser and hotter than it is now. By tuning receivers to different frequencies of microwaves, we can measure the spectrum (the distribution of power arranged by frequency) of this radiation. We find a spectrum that is characteristic of radiation from a body at a temperature of 2.7 degrees above absolute zero. This microwave radiation is not much good for defrosting frozen pizza, but the fact that the spectrum agrees so exactly with that of radiation from a body at 2.7 degrees tells us that the radiation must have come from regions that are opaque to microwaves.

Thus we can conclude that our past light cone must pass through a certain amount of matter as one follows it back. This amount of matter is enough to curve space-time, so the light rays in our past light cone are bent back toward each other.

As one goes back in time, the cross sections of our past light cone reach a maximum size and begin to get smaller again. Our past is pear-shaped.

As one follows our past light cone back still further, the positive energy density of matter causes the light rays to bend toward each other more strongly. The cross section of the light cone will shrink to zero size in a finite time. This means that all the matter inside our past light cone is trapped in a region whose boundary shrinks to zero. It is therefore not very surprising that Penrose and I could prove that in the mathematical model of general relativ ity, time must have a beginning in what is called the big bang. Similar arguments show that time would have an end, when stars or galaxies collapse under their own gravity to form black holes. We had sidestepped Kant's antimony of pure reason by dropping his implicit assumption that time had a meaning independent of the universe. Our paper, proving time had a beginning, won the second prize in the competition sponsored by the Gravity Research Foundation in 1968, and Roger and I shared the princely sum of $300. I don't think the other prize essays that year have shown much enduring value.

There were various reactions to our work. It upset many physicists, but it delighted those religious leaders who believed in an act of creation, for here was scientific proof. Meanwhile, Lifshitz and Khalatnikov were in an awkward position. They couldn't argue with the mathematical theorems that we had proved, but under the Soviet system they couldn't admit they had been wrong and Western science had been right. However, they saved the situation by finding a more general family of solutions with a singularity, which weren't special in the way their previous solutions had been. This enabled them to claim singularities, and the beginning or end of time, as a Soviet discovery.

Most physicists still instinctively disliked the idea of time having a beginning or end. They therefore pointed out that the mathematical model might not be expected to be a good description of space-time near a singularity. The reason is that general relativity, which describes the gravitational force, is a classical theory, as noted in chapter 1, and does not incorporate the uncertainty of quantum theory that governs all other forces we know.

This inconsistency does not matter in most of the universe most of the time, because the scale on which space-time is curved is very large and the scale on which quantum effects are important is very small. But near a singularity, the two scales would be comparable, and quantum gravitational effects would be important. So what the singularity theorems of Penrose and myself really established is that our classical region of space-time is bounded to the past, and possibly to the future, by regions in which quantum gravity is important. To understand the origin and fate of the universe, we need a quantum theory of gravity, and this will be the subject of most of this book.

Quantum theories of systems such as atoms, with a finite number of particles, were formulated in the 1920s, by Heisenberg, Schrödinger, and Dir ac. (Dirac was another previous holder of my chair in Cambridge, but it still wasn't motorized.) However, people encountered difficulties when they tried to extend quantum ideas to the Maxwell field, which describes electricity, magnetism, and light.

One can think of the Maxwell field as being made up of waves of different wavelengths (the distance between one wave crest and the next). In a wave, the field will swing from one value to another like a pendulum.

According to quantum theory, the ground state, or lowest energy state, of a pendulum is not just sitting at the lowest energy point, pointing straight down. That would have both a definite position and a definite velocity, zero. This would be a violation of the uncertainty principle, which forbids the precise measurement of both position and velocity at the same time.

The uncertainty in the position multiplied by the uncertainty in the momentum must be greater than a certain quantity, known as Planck's constant - a number that is too long to keep writing down, so we use a symbol for it:

So the ground state, or lowest energy state, of a pendulum does not have zero energy, as one might expect. Instead, even in its ground state a pendulum or any oscillating system must have a certain minimum amount of what are called zero point fluctuations. These mean that the pendulum won't necessarily be pointing straight down but will also have a probability of being found at a small angle to the vertical.

Similarly, even in the vacuum or lowest energy state, the waves in the Maxwell field won't be exactly zero but can have small sizes. The higher the frequency (the number of swings per minute) of the pendulum or wave, the higher the energy of the ground state.

Calculations of the ground state fluctuations in the Maxwell and electron fields made the apparent mass and charge of the electron infinite, which is not what observations show. However, in the 1940s the physicists Richard Feynman, Julian Schwinger, and Shinichiro Tomonaga developed a consistent way of removing or "subtracting out" these infinities and dealing only with the finite observed values of the mass and charge. Nevertheless, the ground state fluctuations still caused small effects that could be measured and that agreed well with experiment. Similar subtraction schemes for removing infinities worked for the Yang-Mills field in the theory put forward by Chen Ning Yang and Robert Mills. Yang-Mills theory is an extension of Maxwell theory that describes interactions in two other forces called the weak and strong nuclear forces. However, ground state fluctuations have a much more serious effect in a quantum theory of gravity. Again, each wavelength would have a ground state energy. Since there is no limit to how short the wavelengths of the Maxwell field can be, there are an infinite number of different wavelengths in any region of space-time and an infinite amount of ground state energy.

Because energy density is, like matter, a source of gravity, this infinite energy density ought to mean there is enough gravitational attraction in the universe to curl space-time into a single point, which obviously hasn't happened.

One might hope to solve the problem of this seeming contradiction between observation and theory by saying that the ground state fluctuations have no gravitational effect, but this would not work. One can detect the energy of ground state fluctuations by the Casimir effect. If you place a pair of metal plates parallel to each other and close together, the effect of the plates is to reduce slightly the number of wavelengths that fit between the plates relative to the number outside. This means that the energy density of ground state fluctuations between the plates, although still infinite, is less than the energy density outside by a finite amount. This difference in energy density gives rise to a force pulling the plates together, and this force has been observed experimentally.

Forces are a source of gravity in general relativity, just as matter is, so it would not be consistent to ignore the gravitational effect of this energy difference.

Another possible solution to the problem might be to suppose there was a cosmological constant such as Einstein introduced in an attempt to have a static model of the universe. If this constant had an infinite negative value, it could exactly cancel the infinite positive value of the ground state energies in free space, but this cosmological constant seems very ad hoc, and it would have to be tuned to extraordinary accuracy.

Fortunately, a totally new kind of symmetry was discovered in the 1970s that provides a natural physical mechanism to cancel the infinities arising from ground state fluctuations.

Supersymmetry is a feature of our modern mathematical models that can be described in various ways. One way is to say that space-time has extra dimensions besides the dimensions we experience. These are called Grassmann dimensions, because they are measured in numbers known as Grassmann variables rather than in ordinary real numbers. Ordinary numbers commute; that is, it does not matter in which order you multiply them: 6 times 4 is the same as 4 times 6. But Grassmann variables anticommute: x*y is the same as y*x.

Supersymmetry was first considered for removing infinities in matter fields and Yang-Mills fields in a space-time where both the ordinary number dimensions and the Grassmann dimensions were flat, not curved. But it was natural to extend it to ordinary numbers and Grassmann dimensions that were curved. This led to a number of theories called supergravity, with different amounts of supersymmetry. One consequence of supersymmetry is that every field or particle should have a "superpartner" with a spin that is either 1/2 greater than its own or 1/2 less.

The ground state energies of bosons, fields whose spin is a whole number (0, 1, 2 , etc.), are positive. On the other hand, the ground state energies of fermions, fields whose spin is a half number (1/2, 3/2 , etc.), are negative. Because there are equal numbers of bosons and fermions, the biggest infinities cancel in supergravity theories.

There remained the possibility that there might be smaller but still infinite quantities left over. No one had the patience needed to calculate whether these theories were actually completely finite. It was reckoned it would take a good student two hundred years, and how would you know he hadn't made a mistake on the second page? Still, up to 1985, most people believed that most supersymmetric supergravity theories would be free of infinities.

Then suddenly the fashion changed. People declared there was no reason not to expect infinities in supergravity theories, and this was taken to mean they were fatally flawed as theories. Instead, it was claimed that a theory named supersymmetric string theory was the only way to combine gravity with quantum theory. Strings, like their namesakes in everyday experience, are one-dimensional extended objects. They have only length. Strings in string theory move through a background space-time. Ripples on the string are interpreted as particles.

If the strings have Grassmann dimensions as well as their ordinary number dimensions, the ripples will correspond to bosons and fermions. In this case, the positive and negative ground state energies will cancel so exactly that there will be no infinities even of the smaller sort. Superstrings, it was claimed, were the TOE, the Theory of Everything.

Historians of science in the future will find it interesting to chart the changing tide of opinion among theoretical physicists. For a few years, strings reigned supreme and supergravity was dismissed as just an approximate theory, valid at low energy. The qualification "low energy" was considered particularly damning, even though in this context low energies meant particles with energies of less than a billion billion times those of particles in a TNT explosion.

If supergravity was only a low energy approximation, it could not claim to be the fundamental theory of the universe. Instead, the underlying theory was supposed to be one of five possible superstring theories. But which of the five string theories described our universe? And how could string theory be formulated, beyond the approximation in which strings were pictured as surfaces with one space dimension and one time dimension moving through a flat background space-time? Wouldn't the strings curve the background space-time?

In the years after 1985, it gradually became apparent that string theory wasn't the complete picture. To start with, it was realized that strings are just one member of a wide class of objects that can be extended in more than one dimension. Paul Townsend, who, like me, is a member of the Department of Applied Mathematics and Theoretical Physics at Cambridge, and who did much of the fundamental work on these objects, gave them the name "p-branes". A p-brane has length in p directions. Thus a p=1 brane is a string, a p=2 brane is a surface or membrane, and so on. There seems no reason to favour the p=1 string case over other possible values of p. Instead, we should adopt the principle of p-brane democracy: all p-branes are created equal.

All the p-branes could be found as solutions of the equations of supergravity theories in 10 or 11 dimensions. While 10 or 11 dimensions doesn't sound much like the space-time we experience, the idea was that the other 6 or 7 dimensions are curled up so small that we don't notice them; we are only aware of the remaining 4 large and nearly flat dimensions.

I must say that personally, I have been reluctant to believe in extra dimensions. But as I am a positivist, the question "Do extra dimensions really exist?" has no meaning. All one can ask is whether mathematical models with extra dimensions provide a good description of the universe. We do not yet have any observations that require extra dimensions for their explanation. However, there is a possibility we may observe them in the Large Hadron Collider in Geneva. But what has convinced many people, including myself, that one should take models with extra dimensions seriously is that there is a web of unexpected relationships, called dualities, between the models. These dualities show that the models are all essentially equivalent; that is, they are just different aspects of the same underlying theory, which has been given the name M-theory. Not to take this web of dualities as a sign we are on the right track would be a bit like believing that God put fossils into the rocks in order to mislead Darwin about the evolution of life.

These dualities show that the five superstring theories all describe the same physics and that they are also physically equivalent to supergravity. One cannot say that superstrings are more fundamental than supergravity, or vice versa. Rather, they are different expressions of the same underlying theory, each useful for calculations in different kinds of situations. Because string theories don't have any infinities, they are good for calculating what happens when a few high energy particles collide and scatter off each other.

However, they are not of much use for describing how the energy of a very large number of particles curves the universe or forms a bound state, like a black hole. For these situations, one needs supergravity, which is basically Einstein's theory of curved space-time with some extra kinds of matter. It is this picture that I shall mainly use in what follows.

To describe how quantum theory shapes time and space, it is helpful to introduce the idea of imaginary time. Imaginary time sounds like something from science fiction, but it is a well-defined mathematical concept: time measured in what are called imaginary numbers. One can think of ordinary real numbers such as 1, 2, -3.5, and so on as corresponding to positions on a line stretching from left to right: zero in the middle, positive real numbers on the right, and negative real numbers on the left.

Imaginary numbers can then be represented as corresponding to positions on a vertical line: zero is again in the middle, positive imaginary numbers plotted upward, and negative imaginary numbers plotted downward. Thus imaginary numbers can be thought of as a new kind of number at right angles to ordinary real numbers. Because they are a mathematical construct, they don't need a physical realization; one can't have an imaginary number of oranges or an imaginary credit card bill.

One might think this means that imaginary numbers are just a mathematical game having nothing to do with the real world. From the viewpoint b68 of positivist philosophy, however, one cannot determine what is real. All one can do is find which mathematical models describe the universe we live in. It turns out that a mathematical model involving imaginary time predicts not only effects we have already observed but also effects we have not been able to measure yet nevertheless believe in for other reasons. So what is real and what is imaginary? Is the distinction just in our minds?

Einstein's classical (i.e., non-quantum) general theory of relativity combined real time and the three dimensions of space into a 4-D space-time. But the real time direction was distinguished from the three spatial directions; the world line or history of an observer always increased in the real time direction (that is, time always moved from past to future), but it could increase or decrease in any of the three spatial directions. In other words, one could reverse direction in space, but not in time.

On the other hand, because imaginary time is at right angles to real time, it behaves like a fourth spatial direction. It can theref ore have a much richer range of possibilities than the railroad track of ordinary real time, which can only have a beginning or an end or go around in circles. It is in this imaginary sense that time has a shape.

To see some of the possibilities, consider an imaginary time space-time that is a sphere, like the surface of the Earth. Suppose that imaginary time was degrees of latitude. Then the history of the universe in imaginary time would begin at the South Pole. It would make no sense to ask, "What happened before the beginning?". Such times are simply not defined, any more than there are points south of the South Pole. The South Pole is a perfectly regular point of the Earth's surface, and the same laws hold there as at other points. This suggests that the beginning of the universe in imaginary time can be a regular point of space-time, and that the same laws can hold at the beginning as in the rest of the universe. (The quantum origin and evolution of the universe will be discussed in the next chapter.)

Another possible behaviour is illustrated by taking imaginary time to be degrees of longitude on the Earth. All the lines of longitude meet at the North and South Poles. Thus time stands still there, in the sense that an increase of imaginary time, or of degrees of longitude, leaves one in the same spot. This is very similar to the way that ordinary time appears to stand still on the horizon of a black hole. We have come to recognize that this standing still of real and imaginary time (either both stand still or neither does) means that the space-time has a temperature, as I discovered for black holes.

Not only does a black hole have a temperature, it also behaves as if it has a quantity called entropy. The entropy is a measure of the number of internal states (ways it could be configured on the inside) that the black hole could have without looking any different to an outside observer, who can only observe its mass, rotation, and charge. This black hole entropy is given by a very simple formula I discovered in 1974. It equals the area of the horizon of the black hole: there is one bit of information about the internal state of the black hole for each fundamental unit of area of the horizon. This shows that there is a deep connection between quantum gravity and thermodynamics, the science of heat (which includes the study of entropy). It also suggests that quantum gravity may exhibit what is called holography.

Information about the quantum states in a region of space-time may be somehow coded on the boundary of the region, which has two dimensions less. This is like the way that a hologram carries a 3-D image on a 2-D surface. If quantum gravity incorporates the holographic principle, it may mean that we can keep track of what is inside black holes. This is essential if we are to be able to predict the radiation that comes out of black holes. If we can't do that, we won't be able to predict the future as fully as we thought. This is discussed in chapter 4. Holography is discussed again in chapter 7. It seems we may live on a 3-brane - a 4-D (three space plus one time) surface that is the boundary of a 5-D region, with the remaining dimensions curled up very small. The state of the world on a brane encodes what is happening in the 5-D region.

Chapter 3

The Universe in a Nutshell

3.1 *The universe has multiple histories, each of which is determined by a tiny nut*

"*I could be bounded in a nutshell and count myself as king of infinite space*"
(Shakespeare - Hamlet, act 2, scene 2)

Hamlet may have meant that although we, human beings, are very limited physically, our minds are free to explore the whole Universe, and go boldly where even Star Trek fears to tread, bad dreams permitting (?). Is the universe actually infinite, or just very large? And is it everlasting, or just long lived? How could our finite minds comprehend an infinite universe? Isn't it presumptuous of us even to make the attempt? Do we risk the fate of Prometheus, who in classical mythology stole fire from Zeus for human beings to use, and was punished for his temerarity by being chained to a rock, where an eagle picked at his liver?

Despite this cautionary tale, I believe we can and should try to understand the universe. We've already made a remarkable progress in understanding the cosmos, particularly in the last few years. We don't yet have a complete picture, but this may not be far off. The most obvious thing about space is that it goes on and on and on. This has been confirmed by modern instruments, such as the HST, which allows us to probe deep into space. What we see is billions and billions of galaxies, of various shapes and sizes. Each galaxy contains uncounted billions of stars, many of which have planets around them.

We live on a planet orbiting a star in an outer arm of the spiral Milky Way galaxy; the dust in the spiral arms blocks our view of the universe in the plane of the galaxy, but we have a clear line of sight, in cones of directions on each side of the plane, and we can plot the positions of distant galaxies; we find that the galaxies are distributed roughly uniformly throughout space, with some concentrations and voids. The concentration of galaxies appears to drop off at very large distances. But that seems to be because they're so far away and faint that we can't make them out.

As far as we can tell, the universe goes on in space forever. Although the universe seems to be much the same at each point in space, it is definitely changing in time, and this wasn't realized until the early years of the 20-th century; up to then it was thought that the universe was essentially constant in time; it might have existed for an infinite time, but this seems to lead to some absurd conclusions: if stars had been radiating for an infinite time, they would have heated up the universe to their own temperature; even at night, the whole sky would be as bright as the Sun, because every line of sight would lie either on a star or on a cloud of gas that had been heated up until it was as hot as a star.

The observation that we've all made, that the sky at night is dark is very important; it implies that the universe can not have existed forever in the state we see today; something must have happened in the past, to make the stars light up a finite time ago, which means that the light from the very distant stars has not had time to reach us yet. This would explain why the sky at night isn't glowing in every direction. If the stars have just been sitting there forever, why did they suddenly light up a few billion years ago? What was the clock that told them it was time to shine?

As we've seen, this puzzled those philosophers, much like Emanuel Kant, who believed that the universe had existed forever. But for most people it was consistent with the idea that the universe had been created much as it is now only a few thousand years ago. However, discrepancies with this idea began to appear with the observations by Slipher Melvin Vesto and Edwin Hubble, in the 2-nd decade of the 20-th century. In 1923, Hubble discovered that many faint patches of light, called nebulae, were in fact other galaxies - vast collections of stars like our Sun, but at a great distance. In order for them to appear so small and faint, the distances had to be so great that light from them would have taken millions or even billions of years to reach us. This indicated that the beginning of the universe couldn't have been just a few thousand years ago.

3.2 *Chronology of discoveries made by Slipher and Hubble between 1910-1930*:

.1912 - Slipher measured light from 4 nebulae, finding 3 of them redshifted, but Andromeda - blueshifted; his interpretation was that Andromeda was moving towards us, while the other nebulae moved away from us.

.1912-1914 - Slipher measured 12 more galaxies: all, except one, were redshifted.

.1914 - Slipher presented his findings to the American Astronomical Society; Hubble heard the presentation.

.1918 - Hubble began to investigate the nebulae.

.1923 - Hubble determined that the spiral nebulae, including Andromeda, are other galaxies.

.1914-1925 - Slipher and others kept measuring doppler shifts; the score in 1925 was 43 redshifts to 2 blueshifts.

.1929 - Hubble and Milton Humason, after continuing to measure Doppler shifts and finding that on the large scale, every galaxy appears to be receding from every other, announced the discovery that the universe in expanding.

But the second thing Hubble discovered was even more remarkable; astronomers have learned that by analysing light from other galaxies, it was possible to measure whether they're moving towards us or away from us; to their great surprise, they found that nearly all galaxies are moving away; moreover, the farther they are from us, the faster they're moving away.

It was Hubble who recognized the dramatic implications of this discovery; on a large scale, every galaxy is moving away from every other galaxy; the universe is expanding.

3.3 *Hubble's Law*

By analysing the light from other galaxies, Edwin Hubble discovered in the 1920's that nearly all galaxies are moving away from us with a velocity V that is proportional to their distance R from the Earth; so $V=H*R$; this important observation known as Hubble's law, established that the universe is expanding, with the Hubble's constant H setting the rate of expansion.

3.4 *The Doppler Effect*

The relationship between speed and wavelength, which is called the Doppler effect, is an every day experience; listen to a plane that passes over head; as it approaches, its engines sound at a higher pitch, and when it passes and disappears, it sounds at a lower pitch; the higher pitch corresponds to waves with a shorter wavelength (the distance between one crest the next), and the higher the frequency, the higher the number of crests per second; this is because, as the plane moves toward you, it would be nearer to you when it emits the next wave crest, lessening the distance between wave crests; similarly, as the plane moves away, the wavelengths increase, and the pitch you perceive is lower.

The Doppler effect is also true of light waves; if a galaxy were to remain at a constant distance from earth, characteristic lines in the spectrum would appear in a normal or standard position; however, if the galaxy is moving away from us, the waves will appear elongated, or stretched, and if the galaxy is moving toward us, then the waves will appear to be compressed.

The discovery of the expansion of the universe was one of the great intellectual revolutions of the 20-th century; it came as a total surprise , and it completely changed the discussion of the origin of the universe. If the galaxies are moving apart, they must have been closer together in the past. From the present rate of expansion we can estimate that they must have been very close together indeed, 10 to 15 billion years ago.

As described in chapter 2, Roger Penrose and I were able to show that Einstein's general theory of relativity implied that the universe, and time itself, must have had a beginning in a tremendous explosion. Here was the explanation of "why the sky at night was dark". No star could have been shining for longer than $(10 - 15)*10^9$ years, the time since the Big Bang.

3.5 *The hot Big Bang*

If general theory of relativity were correct, the universe started with an infinite temperature and density, at the Big Bang singularity. As the universe expanded, the temperature of radiation decreased; at about 1/100 of a second after the Big Bang, the temperature would have been $100*10^9$ degrees, and the universe would have contained mostly photons, electrons and neutrinos - extremely light particles - and their antiparticles, together with some protons and neutrons. For the next 3 minutes, as the universe cooled to about 10^9 degrees, protons and neutrons would have combined to form nuclei of Hydrogen, Helium and other light elements; hundreds of thousands of years later, when the temperature would have dropped to few thousands degrees, the electrons would have slowed down to the point when the light nuclei could capture them to form atoms. However, the heavier elements, of which we are made, such as Carbon and Oxygen, would not form until billions of years later, from the burning of He in the centre of stars. This picture of a dense, hot, early stage of the universe was first put forward by the scientist George Gamow, in 1948, in a paper he wrote with Ralph Alpher, which made the remarkable prediction that the radiation from this very hot early stage, should still be around us today. Their prediction was confirmed in 1965, when the physicists Arno Penzias and Robert Wilson observed the CMBR.

We're used to the idea that events are caused by earlier events, which in turn, are caused by still earlier events.

There's a chain of causality, stretching back into the past; but suppose this chain has a beginning, suppose there was a first event; what caused it?

This was not a question many scientists wanted to address. They tried avoid it, either by claiming, like the Russians that the universe didn't have a beginning, or by maintaining that the origin of the universe didn't lie within the realm of science, but belonged to metaphysics or religion.

In my opinion this is not a position any true scientist should take; if the laws of science are suspended at the beginning of the universe, might not they fail at other times also? A law is not a law if it only holds sometimes. We must try to understand the beginning of the universe on the basis of science. It may be a task beyond our powers, but we should at least make the attempt.

While the theorems that Penrose and I proved, that showed that the universe must have had a beginning, they didn't give much information about the nature of that beginning. They indicated that the universe began in a big bang, a point when the whole universe, and everything in it, was scrunched up into a single point of infinite density. At this point, Einstein's general theory of relativity would have broken down, so it can't be used to predict in what manner the universe began. One is left with the origins of the universe apparently beyond the scope of science; this was not a conclusion that scientists should be happy with. As chapters 1 and 2 point out, the reason general theory of relativity broke down near the big bang is that it didn't incorporate the uncertainty principle, the random element of quantum theory, that Einstein had objected to, on the ground that "God does not play dice". However, all the evidence is that God is quite a gambler.

One can think of the universe as being like a giant casino, with dice being rolled and wheels being spun on every occasion. You might think that operating a casino is a very chancy business, because you risk losing money every time dice are thrown or wheels are spun, but over a large number of bets, the gains and losses average out to a result that can be predic ted; eventhough the result of any particular bet cannot be predicted. The casino operators make sure that the odds average out in their favour; that's why the casino operators are so rich; the only chance you have of winning against them is to stake all of your money on a few rolls of the dice or spins of the wheel.

It's the same with the universe; when the universe is big as it is today, there are very large numbers of rolls of the dice, and the results average out to something one could predict; that's why classical laws work for large systems. But when the universe was very small, as it was near in time for the big bang, there are only a small number of rolls of the dice, and the uncertainty principle becomes (is) very important.

Because the universe keeps on rolling the dice to see what happens next, it doesn't have just a single history, as one might have thought, instead, the universe must have every possible history, each with its own probability. There must be a history of the universe in which Belize won every gold medal at the Olympic Games, though maybe the probability is low.

This idea that the universe has multiple histories may sound like SF but it is now accepted as a science fact. It was formulated by Richard Feynman, who was both a great physicist and quite a character.

3.6 *Feynman stories*

Born in Brooklyn - New York, in 1918, Richard Feynman completed his PhD under John Wheeler, at Princeton University in 1942; shortly afterward, he was drawn into the Manhattan project. He was known for both his exuberant personality and practical jokes (at the Los Alamos labs he enjoyed cracking into the top secret safes), and for being an exceptional physicist. He became a key contributor to Atomic Bomb theory. Feynman's perpetual curiosity about the world was the very root of his being; (but is) not only the engine for his scientific success it led him to numerous astonishing exploits, such as deciphering Mayan hieroglyphs.

In the years following the WW2, Feynman found a powerful new way of thinking about quantum mechanics, for which he was awarded the Nobel Prize in 1965. He challenged the basic classical assumption that each particle has one particular history, instead he suggested that particles travel from one location to another along every possible path through space-time; with each trajectory, Feynman associated two numbers - one for the size (the amplitude) of a wave, and one for the phase (whether it's at a crest or a trough). The probability that a particle going from A to B is found by adding up the waves associated with every possible path that passes through A and B. Nevertheless, in the every day world it seems to us that objects follow a single path between their origin and destination point. This agrees with Feynman's multiple history idea, because, for large objects, his rule for assigning numbers to each path ensures that all paths, but one, cancel out when their contributions are combined. Only one of the infinity of paths matters, as far as the motion of macroscopic objects is concerned, and this trajectory is precisely the one emerging from Newton's classical laws of motion.

We're now working to combine Einstein's general theory of relativity and Feynman's idea of multiple histories into a complete unified theory that would describe everything that happens in the universe.

This unified theory will enable us to calculate how the universe would develop, if we know how the history started. But the unified theory will not in itself tell us how the universe began, or what its initial state was. For that we need what are called "boundary conditions" - rules that tell us what happened at the frontiers of the universe, the edges of space and time. If the frontier of the universe was just at a normal point of space and time, we can go past it and claim the territory beyond as part of the universe. On the other hand, if the boundary of the universe was at a jugged edge, where space and time were scrunched up, and the density was infinite, it would be very difficult to define boundary conditions.

However, a colleague (named Jim Hartland) and I realised there was a third possibility: maybe the universe has no boundary in space and time. At first sight this seems to be in direct contradiction with the theorems that Penrose and I proved, which showed that the universe must have a beginning, a boundary in time.

However, as explained in chapter 2, there is another kind of time called imaginary time, that is at right angles to the ordinary real time that we feel going by. The history of the universe in real time determines its history in imaginary time, and vice versa; but the two kinds of histories can be very different; in particular, the universe need have no beginning or end in imaginary time. (Imaginary time behaves just like another direction in space). Thus, the histories of the universe in imaginary time can be thought of as curved surfaces, like a ball, a plane or a saddle shape, but with 4 dimensions instead of 2. If the histories of the universe went off to infinity, like a saddle or a plane, one would have the problem of specifying what the boundary conditions were at infinity.

But one would avoid having to specify boundary conditions at all, if the histories of the universe in imaginary time are closed surfaces, like the surface of the Earth; the surface of the Earth doesn't have any boundary or edges (there are no reliable reports of people falling off). If the histories of the universe are indeed closed surfaces, as Hartland and I proposed, it would have fundamental implications for philosophy, and our picture of where we came from; the universe would be entirely self-contained; it wouldn't need anything outside to wind up the clock-work and set it going; instead, everything in the universe would be determined by the laws of science and by rolls of the dice within the universe.

This may sound presumptuous, but it's what I and many other scientists believe; even if the boundary condition of the universe is that it has no boundary, it won't have just a single history, it would have multiple histories, as suggested by Feynman. There would be a history in imaginary time corresponding to every possible closed surface; and each history in imaginary time would determine a history in real time. Thus we have a super abundance of possibilities for the universe. What picks up the particular universe that we live in from the set of all possible universes? One point we can notice is that many of the possible histories won't go through the sequence of forming galaxies and stars, that was essential to our own development. While it may be that intelligent beings can evolve without galaxies and stars, this seems unlikely. Thus, the very fact that we exist as beings who can ask the question "why is the universe the way it is ?" is a restriction on the history we live in. It implies it is one of the minority of histories that has galaxies and stars.

This is an example of what is called "the anthropic principle".

3.7 *The anthropic principle*

Roughly speaking, the anthropic principle says that we see the universe the way it is, at least in part, because we exist. It is a perspective that is diametrically opposed to the dream of a fully predictive unified theory, in which the laws of nature are complete, and the world is the way it is because it couldn't be otherwise.

There are a number of different versions of the anthropic principle, ranging from those that are so weak as to be trivial, to those that are so strong as to be absurd. Although most scientists are reluctant to adopt the strong version of the anthropic principle, a few people would quirrel with the utility of some weak anthropic principle arguments. The weak anthropic principle amounts to an explanation of which of the various possible eras or parts of the universe we would inhabit. For instance, the reason why the big bang occurred, about 10^{10} years ago, is that the universe must be old enough, so that some stars will have completed their evolution, to produce elements like O and C, out of which we are made, and young enough, so that some stars would still be producing energy to sustain life.

Within the framework of the no boundary proposal, one can use Feynman's rule for assigning numbers to each history of the universe, to find which properties of the universe are likely to occur. In this context, the anthropic principle is implemented by requiring that the histories contain intelligent life. one would feel happier about the anthropic principle of course, if one could show that a number of different initial configurations for the universe are likely to have evolved, to produce a universe like the one we observe. This would imply that the initial state of the part of the universe that we inhabit did not have to be chosen with great care.

The anthropic principle says that the universe has to be more or less as we see it, because if it were different, there wouldn't be anyone here to observe it. Many scientists dislike the anthropic principle, because it seems rather vague and doesn't appear to have much predictive power, but the anthropic principle can be given a precise formulation, and it seems to be essential when dealing with the origin of the universe.

M-theory, described in chapter 2, allows a very large number of possible histories for the universe; most of these histories are not suitable for the development of intelligent life: either they're empty, last for too shorted time, are too highly curved, or wrong in some other way. Yet, according to Richard Feynman's idea of multiple histories, these uninhabited histories can have quite a high probability; in fact, it doesn't really matter how many histories there may be that don't contain intelligent beings; we're interested only in a subset of histories in which intelligent life develops.

This intelligent life needn't be anything like humans, little green aliens would do as well - they might do rather better; the human race doesn't have a very good record of intelligent behaviour. As an example of the power of the anthropic principle, consider the number of directions in space; it's a matter of common experience that we live in a 3-D space. That's to say that we can represent the position of a point in space by three numbers: for example (latitude, longitude, height above sea level). But why is space three dimensional? Why isn't it 2 or 4 or some other number of dimensions, as in SF-s. In M-theory space has 9 or 10 dimensions, but it is thought that 6 or 7 of the directions are curled up very small, leaving 3 dimensions that are large and nearly flat. Why don't we live in a history in which 8 of the dimensions are curled up small, leaving only 2 dimensions that we notice?

A 2-D animal would have a hard job digesting food; if it had a gut that went right through it, it would divide the animal in two, and the poor creature would fall apart. So, two flat directions are not enough for anything as complicated as intelligent life.

On the other hand, if there were four ore more nearly flat directions, the gravitational force between two bodies would increase more rapidly as they approached each other. This would mean that planets would not have stable orbits about their suns; they would either fall into their suns or escape to the outer darkness and cold. Similarly, the orbits of electrons in atoms would not be stable, so matter as we know it would not exist.

Thus, although the idea of multiple histories would allow any number of nearly flat directions, only histories with three flat directions will contain intelligent beings. Only in such histories would the question be asked "why the space has three dimensions?"

The simplest history of the universe in imaginary time is a round sphere, like the surface of the Earth, but with 2 more dimensions. It determines a history of the universe in the real time the we experience, in which the universe is the same at every point of space, and is expanding in time. In these respects, it is like the universe we live in, but the rate of expansion is very rapid, and it keeps on getting faster. Such accelerated expansion is called inflation, because it's like the way prices go up and up at an ever increasing rate.

3.8 ***Inflation may be a law of nature***

Inflation in Germany rose after the peace until, by February 1920, the price level was as 5 times as high as it has been in 1918. After July 1922, the phase of hyper-inflation began; all confidence in money vanished, and the price index rose faster and faster, for 15 months, outpacing the printing process, which couldn't produce money as fast as it was depreciating. By late 1923, 300 paper mills were working at top speed, and 150 printing companies had 200 presses running day and night, turning out currency.

Inflation in prices is generally held to be a bad thing, but in the case of the universe, inflation is very beneficial. The large amount of expansion smoothes out any lumps and bumps that there may have been in the early universe. As the universe expands, it borrows energy from the gravitational field to create more matter. The positive matter energy is exactly balanced by the negative gravitational energy, so the total energy is zero. When the universe doubles its size, the matter and gravitational energies both double; so twice zero is still zero. If only the banking world was so simple.

If the history of the universe in imaginary time were a perfectly round sphere, the corresponding history in real time would be an universe that continued to expand in an inflationary manner forever. While the universe is inflating, matter could not fall together to form galaxies and stars, and life - let alone intelligent life, like us - couldn't develop; thus, although histories of the universe in imaginary time, that are perfectly round spheres, are allowed by the notion of multiple histories, they're not of much interest. However, histories in imaginary time that are more flattened at the South Pole are much more relevant.

In this case, the corresponding history in real time will expand in an accelerated inflationary manner at first, but then, the expansion would begin to slow down, and galaxies could form. In order for intelligent life to be able to develop, the flattening at the South Pole must be very slight; this would mean that the universe will expand initially by an enormous amount. The record level of monetary inflation occurred in Germany between WWs, when prices rose billions of times; but the amount of inflation that must have occurred in the universe is at least a billion billion billion times that.

Because of the uncertainty principle, there won't be just one history of the universe that contains intelligent life, instead, the histories in imaginary time will be a whole family of slightly deformed spheres, each of which corresponding to a history in real time in which the universe inflates for a long time, but not indefinitely. We can then ask which of these allowable histories is the most probable.

It turns out that the most probable histories are not completely smooth, but have tiny ups and downs. The ripples in the most probable histories really are minuscule; the departures from smoothness are of the order of 1 part in 10^5. Nevertheless, although they are extremely small, we've managed to observe them as small variations in the microwaves that come to us from different directions in space.

The COBE satellite was launched in 1989 and made a map of the sky in microwaves. Different colours indicate different temperatures, but the whole range from red to blue is only about 10^{-4} of a degree; yet this is enough variation between different regions of the early universe for the extra gravitational attraction in the denser regions to stop them expanding eventually, and to cause them to collapse again under their own gravity, to form galaxies and stars.

So, in principle at least, the COBE map is the blueprint of all the structures in the universe.

What will be the future behaviour of the most probable histories of the universe that are compatible with the appearance of intelligent life (beings)? There seems to be various possibilities, depending on the amount of matter in the universe. If there's more than a certain critical amount, the gravitational attraction between galaxies will slow them down and will eventually stop them from flying apart; they would then start falling towards each other, and will all come together in a Big Crunch, that would be the end of the history of the universe in real time.

If the density of the universe is bellow the critical value, gravity is too weak to stop the galaxies from flying apart forever. All the stars will burn out and the universe will get increasingly emptier and colder; so, again, things will come to an end, but in a less dramatic way.

Either way, the universe will last a few billion years more. As well as matter, the universe may contain what is called vacuum energy - energy that is present even in apparently empty regions of space. By Einstein's famous equation $E = mc^2$, this vacuum energy has mass; this means that it has a gravitational effect on the expansion of the universe; but remarkably enough, the effect of vacuum energy is the opposite of that of matter; matter causes the expansion to slow down, and can eventually stop and reverse it; on the other hand, vacuum energy causes the expansion to accelerate, as in inflation. In fact vacuum energy acts just like the cosmological constant mentioned in chapter 1, that Einstein added to his original equations in 1917, and when he realised that they didn't admit a solution representing a static universe.

After Hubble's discovery of the expansion of the universe, this motivation for adding a term to the equations disappeared, and Einstein rejected the cosmological constant as a mistake.

However, it may not have been a mistake at all; as described in chapter 2, we now realise that quantum theory implies that space-time is filled with quantum fluctuations; in a supersymmetric theory, the infinite positive and negative energies of these ground state fluctuations, cancel out between particles of different spin. But we wouldn't expect the positive and negative energies to cancel so completely that there wasn't a small finite amount of vacuum energy left over, because the universe is not in a supersymetric state.

The only surprise is that the vacuum energy is so nearly zero that it was not obvious some time ago. Maybe this is another example of the anthropic principle.

A history with a larger vacuum energy could not have formed galaxies, so would not contain beings who could ask the question "why is the vacuum energy so low". We can try to determine the amounts of matter and vacuum energy in the universe from various observations.

We can show the results in a diagram (X = matter density, Y = vacuum energy); a dotted line shows the boundary of the region in which intelligent life could develop. Observations of super-nova clustering and the microwave background, each mark out regions in this diagram. Fortunately, all regions have a common intersection.

If the matter density and vacuum energy lie in this intersection, it means that the expansion of the universe has begun to speed up again, after a long period of slowing down. It seems that inflation may be a law of nature.

In this chapter we've seen how the behaviour of the vast universe can be understood in terms of history in imaginary time, which is a tiny, slightly flattened sphere; it's like Hamlet's nutshell, yet this nut encodes everything that happens in real time.

So, Hamlet was quite right; we could be bounded in a nutshell and still count ourselves kings of infinite space.

Chapter 4

Predicting the Future

4.1 ***How the loss of information in black holes may reduce our ability to predict the future***

The human race has always wanted to control the future, or at least to predict what will happen; that's why astrology is so popular. Astrology claims that events on earth are related to the motions of the planets across the sky.; this is a scientifically testable hypothesis, or would be if astrologers stuck their necks out and made definite predictions that could be tested. However, wisely enough they make their forecast so vague that they could apply to any outcome; statements such as "personal relations may become intense" or "you will have a financially rewarded opportunity" can never be proved wrong. But the real reason most scientists don't believe in astrology is not scientific evidence (or the lack of it), but because it's not consistent with other theories that have been tested by experiments.

When Copernicus and Galileo discovered that the planets orbit the Sun rather than the Earth, and Newton discovered the laws that govern their motion, astrology became extremely implausible.
Why should the position of other planets against the background sky, as seen from earth, have any correlation with the macromolecules on a minor planet, that call themselves intelligent life; yet this is what astrology would have us believe. There is no more experimental evidence for some of the theories described in this book than there is for astrology, but we believe them because they're consistent with theories that have survived testing. The success of Newton's laws and other physical theories led to the idea of scientific determinism, which was first expressed at the beginning of the 19-th century by the French scientist Le Marquis de LaPlace. LaPlace suggested that if we knew the positions and velocities of all the particles in the universe at one time, the laws of physics should allow us to predict what the state of the universe would be at any other time, in the past or in the future. In other words, if scientific determinism holds, we should, in principle, be able to predict the future, and wouldn't need astrology.

Of course, in practice, even something as simple as Newton's theory of gravity produces equations that we can't solve exactly for more than two particles; furthermore, the equations often have a property known as chaos, so that a small change in position or velocity at one time, can lead to completely different behavior at later times; as those who have seen "Jurassic Park" know, a tiny disturbance in one place can cause a major change in another: a butterfly flapping his wings in Tokyo can cause rain in New York Central Park; the trouble is: the sequence of events is not repeatable, so the next time the butterfly flaps his wings, a host of other factors would be different, and would also influence the weather; that's why weather forecasts are so unreliable.

Thus, although in principle, the laws of quantum electrodynamics should allow us to calculate everything in chemistry and biology, we've not had success in predicting human behavior from mathematical equations. Nevertheless, despite these practical difficulties, most scientists have comforted themselves with the idea that, again, in principle, the future is predictable.

At first sight, determinism would also seem to be threatened by the uncertainty principle, which says that we can't measure accurately both the position and velocity of a particle at the same time; the more accurately we measure the position, the less accurate we can determine the velocity, and vice-versa. The LaPlace version of scientific determinism held that if we knew the position and velocity of particles at one time, we could determine their position and velocity at any time in the past or future. But how can we even get started, if the uncertainty principle prevents us from knowing accurately both the position and velocity at one time? However good our computer is, if you put lousy data in, we get lousy predictions out. However, determinism was restored in a modified form, in a new theory called quantum mechanics, which incorporated the uncertainty principle; in quantum mechanics, one can, roughly speaking, accurately predict half what one would expect to predict in classical LaPlace point of view. In quantum mechanics a particle doesn't have a well defined position or velocity, but its state can be represented by what is called a wave function.

A wave function is a number at each point of space that gives the probability that a particle is to be found at that position. The rate at which the wave function changes from point to point tells how probable different particle velocities are. Some wave functions are sharply peaked at a particular point in space; in these cases there is only a small amount of uncertainty in the position of a particle; but we can also see in a diagram that in such cases the wave function changes rapidly near the point: up on one side and down on the other. That means the probability distribution for the velocity is spread over a wide range; in other words, the uncertainty in the velocity is large. Consider, on the other hand, a continuous train of waves; now there is a large uncertainty in position but a small uncertainty in velocity; so the description of a particle by a wave function does not have a well defined position or velocity; it satisfies the uncertainty principle. We now realize that the wave function is all that could be well defined. We can't even suppose that a particle has a position and velocity that are known to God, but are hidden from us. Such Him -variable theories product results that are not in agreement with observation. Even God is bound by the uncertainty principle and can't know the position and velocity; He can only know the wave function. The rate at which the wave function changes with time is given by what is called "the Schroedinger equation". If we know the wave function at one time, we can use the Schroedinger equation to calculate it at any other time, past or future. Therefore there is still determinism in quantum theory, but it's on a reduced scale. Instead of being able to predict both the position and velocity we can predict only the wave function. This can allow us to predict either the position or velocity, but not both accurately. Thus, in quantum theory, the ability to make exact predictions is just half of what was in the classical LaPlace world view. Nevertheless, within this restricted sense, it is still possible to claim that there is determinism. However, the use of the Schroedinger equations to evolve a wave function forward in time, that is to predict what it will be at future times, implicitly assume that time runs on smoothly, everywhere, forever. This was certainly true in Newtonian physics; time was assumed to be absolute, meaning that each event in the history of the universe was labeled by a number called time, and that a series of time labels ran smoothly from the infinite past to the infinite future. This is what might be called the "common sense view of time" and it is the view of time that most people and even most physicists have at the back of their minds.

However, in 1905, as we've seen, the concept of absolute time was overthrown by the special theory of relativity, in which time was no longer an independent quantity on its own, but was just one direction in a 4-D continuum, called space-time. In special relativity different observers, traveling at different velocities, move through space-time on different paths; each observer has his own measure of time along the path he's following; and different observers will measure different intervals of time between events; thus in special relativity, there is no unique, absolute time that we can use to label events.

However, the space-time of special relativity is flat; this means that in special relativity the time measured by any freely moving observer increases smoothly in space-time from (-8) in the infinite past to (+8) in the infinite future. We can use any of these measures of time in the Schroedinger equation to evolve a wave function in special relativity, therefore, we still have the quantum version of determinism. The situation was different in the general theory of relativity, in which space-time ws not flat but curved and distorted by the matter and energy in it. In our solar system, the curvature of space-time is so slight, at least on a macroscopic scale, that it doesn't interfere with our usual idea of time. In this situation we can still use this time in the Schroedinger equation to get a deterministic evolution of the wave function. However, once we allow space-time to be curved, the door is open to the possibility that it may have a structure that doesn't admit a time that increases smoothly for every observer as we would expect for a reasonable measure of time. For example, suppose that space time was like a vertical cylinder. Height up the cylinder would be a measure of time that increased for every observer and ran from (-8) to (+8). However, imagine instead that space-time was like a cylinder with a handle or wormhole, that branched off and then join back; then any measure of time would necessarily have stagnation points where the handle joined the main cylinder, points where time stood still; at these points time would not increase for any observer; in such a space-time we couldn't use the Schroedinger equation to get a deterministic evolution for the wave function. Watch out for wormholes! You never know what may come out of them. Black holes are the reason we think time will not increase for every observer. The first discussion of black holes appeared in 1783 (?). A former Cambridge Don, John Mitchell, presented the following argument: "if one fires a particle, such as a cannon ball, vertically upward, its ascent will be slowed down by gravity, and eventually, the particle would stop moving upward and will fall back. However, if the initial upward velocity is greater than a critical value called the escape velocity, gravity will never be strong enough to stop the particle, and it will get away." The escape velocity is about 12km/s for the earth and about 100km/s for the Sun; both of these escape velocities are much higher than the speed of real cannon balls, but they're small compared to the speed of light. Thus light can get away from the earth and the sun without much difficulty. However, Mitchell argued that there could be stars that could be much massive than the sun, and have escape velocities greater than the speed of light. We would not be able to see these stars, because any light they sent out would be dragged back by the gravity of the star. Thus, they would be what Mitchell called dark stars, and we now call black holes. Mitchell's idea of dark stars was based on Newtonian physics, in which time was absolute, and went on regardless of what happened; thus, they didn't affect our ability to predict the future in the classical Newtonian picture. But the situation was very different in the general theory of relativity in which massive bodies curve space-time. In 1916, shortly after the theory was first formulated, Karl Schwartzschild (who died soon after by an illness contracted on the Russian front in WW1) found a solution for the field equations of general relativity, that represented a black hole.

4.2 *The Schwartzschild black hole*

Schwartzschild's work revealed a stunning implication of general relativity; he showed that if the mass of a star is concentrated in a small enough region, the gravitational field at the surface of the star becomes so strong that even light can no longer escape; this is what we now call a black hole, a region in space-time bounded by a so-called event horizon, from which it is impossible for anything, including light, to reach a distant observer. For a long time, most physicists, including Einstein, were skeptical whether such extreme configurations of matter could actually ever occur in the real universe. However, we now understand that when any sufficiently heavy, non-rotating star, however complicated its shape and internal structure, runs out of nuclear fuel, it will necessarily collapse to a perfectly spherical Schwartzschild black hole. What Schwartzschild has found wasn't understood, or its importance recognized for many years. Einstein himself never believed in black holes, and his attitude was shared by most of the old guard in general relativity. I remember going to Paris to give a seminars on my discovery that quantum theory means that black holes aren't completely black. My seminar fell rather flat because at that time almost no one in Paris believed in black holes. The French also felt that the name, as they translated it "trous noirs" had dubious sexual connotations, and should be replaced by "astres occlus", or "hidden stars". However, neither this or other suggested names caught the public's imagination than the term black holes, which was first introduced by John Archibald Wheeler, the American physicist who inspired much of the modern work in this field.

4.3 *John Wheeler*

John Archibald Wheeler, was born in 1911 in Jacksonville - Florida. He earned his PhD from Joan Hopkins University in 1933 for his work on the scattering of light by the Helium atom. In 1938 he worked with the Danish physicist Niels Bohr to develop the theory of nuclear fission. For a while thereafter, Wheeler, together with his graduated student Richard Feynman, concentrated on the study of electrodynamics; but shortly after America entered WW2, both went on to contribute to the Manhattan project. In the early 50's, inspired by Oppenheimer's work in 1939 on the gravitational collapse of the massive stars, Wheeler turned his attention to Einstein's theory of general relativity. At that time most physicists were caught up in the study of nuclear physics, and general relativity was not really regarded as relevant to the physical world. But almost single handedly, Wheeler transformed the field, both through his research and his teaching of Princeton's first course on relativity. Much later, in 1969, he coined the term black hole, for the collapsed state of matter which few yet believed is real; inspired by the work of Verner Israel, he conjectured that "black holes have no hair", which meant that the collapsed state of any massive non-rotating star, could in fact be described by Schroedinger's solution.

The discovery of quasars in 1963 brought forth an outburst of theoretical work on black holes, an observational attempt to detect them. Here is the picture that has emerged: Consider what we believe to be the history of a star with a mass 20*M(Sun); such stars form from clouds of gas, like those in the Orion nebula; as clouds of gas contract under their own gravity, the gas heats up and eventually becomes hot enough to start the nuclear fusion reaction that converts the H into He. The heat generated by this process creates a pressure that supports the star under its own gravity, and stops it form contracting further; a star will stay in this state for a long time, burning H and radiating light into space. The gravitational field of the star will affect the paths of light rays coming from it; one can draw a diagram with time plotted upwards and distance from the center of the star plotted horizontally. In this diagram, the surface of the star is represented by two vertical lines, one on either side of the center. One can choose that time be measured in seconds and distance in light-seconds; when we use this units, the speed of light is 1 (c = 1 Ls/s). This means, that far from the star and its gravitational field, the path of a light ray on the diagram is a line at a 45° angle to the vertical. However, nearer the star, the curvature of space-time produced by the mass of the star would change the path of the light rays and cause them to be at a smaller angle to the vertical. Massive stars convert their H into He much faster than the Sun does; this means they can run out of H in as little as few 10^8 years; after that, such stars face a crisis - they can burn their He into heavier elements, such as C and O, but these nuclear reactions don't release much energy, so the stars loose heat and the thermal pressure that supports them against gravity; therefore they begin to get smaller. If they're more than about twice the mass of our sun, the pressure would never be sufficient to stop the contraction; they'll collapse to a zero size and infinite density, to form what's called a singularity. In a diagram of time against distance from the center, as the star shrinks, the paths of light rays from its surface will start out at smaller and smaller angles from the vertical.

When the star reaches a certain critical radius, the path will be vertical on the diagram, which means that the light will hover at a constant distance from the center of the star, never getting away. This critical path of light will sweep out a surface called the event horizon, which separates the region of space-time from which light can escape from the regions form which it can not. Any light emitted by the star after it passes the event horizon would be bent back inward by the curvature of space-time; this star will have become one of Mitchell's dark stars, or as we say now, a black hole. How can you detect a black hole if no light can get out of it?

The answer is that black holes exert the same gravitational pull on the neighboring objects as did the star that collapsed. If the Sun were a black hole, or had managed to become one without losing any of its mass, the planets would still orbit as they do now. One way of searching for black holes is therefore to look for matter that is orbiting what seems to be an unseen massive compact object. A number of such systems have now been (detected) observed, perhaps the most impressive are the black hole giants that occur in the centers of galaxies and quasars. The properties of black holes that have been discussed thus far don't raise great problems with determinism. Time will come to an end for an astronaut who falls into a black hole and hits the singularity; however, in the general relativity, one is free to measure time at different rates in different places.

One could therefore speed up the astronaut's watch as he approaches the singularity, so that it still registered an infinite interval of time. On the time-distance diagram, the surfaces of constant values of this new time would be all crowded together, in the center, below the point where the singularity appeared. But they would agree with the usual measure of time in the nearly flat space-time, far away from the black hole. One could use this time in the Schroedinger equation and calculate the wave function at later times, if one knew it initially; thus, one still has determinism.

It's worth noting, however, that at late times part of the wave function is inside the black hole, where it can't be observed by someone outside; thus, an observer who is sensible enough not to fall into a black hole, can't write the Schroedinger equation backward and calculate the wave function at early times. To do that, he would need the part of the wave function that's inside the black hole; this contains the information about what fell into the black hole. This is potentially a very large amount of information, because a black hole, of a given mass and rate of rotation, can be formed of a very large number of different collection of particles; a black hole does not depend on the nature of the body that had collapsed to form it.

John Wheeler called this result "a black hole has no hair", but for the French, this just confirmed their suspicions. The difficulty with determinism arose when i discovered that black holes aren't completely black. As we saw in chapter 2, quantum theory means that fields can't be exactly zero, even in what is called the vacuum; if they were zero they would have both an exact value or position at zero, and an exact rate of change or velocity, that was also zero; this would be a violation of the uncertainty principle, which says that the position and velocity can't both be well defined. All fields must instead have a certain amount of what are called "vacuum fluctuations", in the same way that the pendulum in chapter 2 had to have "zero-point fluctuations".

Vacuum fluctuations can be interpreted in several ways that seem different, but are in fact mathematically equivalent. From a positivist viewpoint, one is free to use whatever picture is most useful for the problem in question. In this case is helpful to think of vacuum fluctuations as pairs of virtual particles that appear together at some point of space-time, move apart and come back together, and annihilate each other. Virtual means that these particles can't be observed directly, but their indirect effects can be measured, and they agree with theoretical predictions to a remarkable degree of accuracy. If a black hole is present, one member of the pair of virtual particles may fall into the black hole, leaving the other member free to escape to infinity. To someone far from the black hole the escaping particles appear to have been radiated by the black hole. The spectrum of a black hole is exactly what we would expect from a hot body with the temperature proportional to the gravitational field on the horizon - the boundary of the black hole; in other words, the temperature of a black hole depends on its size. A black hole of a few solar masses would have a temperature of about 10^-6 of a degree above absolute zero, and a larger black hole would have an even lower temperature; thus, any quantum radiation from such black holes would be swamped by the -2.7 degree radiation left over from the hot big bang - the CBR that we discussed in chapter 2. It would be possible to detect radiation from much smaller black holes, but it don't seem to be many of them around. That's a pity; if one were discovered, I would get a Nobel Prize. However, we have indirect observational evidence for this radiation, and that evidence comes from the early universe.

As described in chapter 3, it's thought that very early in its history, the universe went through an inflationary period, during which it expanded at an ever increasing rate. The expansion during this period would have been so rapid that some objects would be too distant from us for their light ever to rach us. The universe would have expanded too much and too rapidly while that light was traveling toward us. Thus, there would be an horizon in the universe, like the horizon of a black hole, separating the region from which light can reach us and the region from which it can't. Very similar arguments show that there should be thermal radiation from his horizon as there is from a black hole's horizon. In thermal radiation we've learned to expect a characteristic spectrum at density fluctuations; in this case, the density fluctuations would have expanded with the universe.

When their length scale became longer than the size of the event horizon, they would have become frozen in, so that we can observe them today as small variations in the temperature of the CBR left over from the early universe. The observation of these variations agree with the prediction of thermal fluctuations with remarkable accuracy. Even if the observational evidence for black hole radiation is a bit indirect, every one who has studied the problem agrees it must occur in order to be consistent with our other observationally tested theories.

This has an important implication for determinism; the radiation from a black hole will carry away energy, which must mean that the black hole will loose mass and get smaller; in tum, this would mean that its temperature will rise and the rate of radiation will increase. Eventually the black hole will get down to zero mass. We don't know how to calculate what happens at this point, but the only natural reasonable outcome would seem to be that the black hole disappears completely. So what happens then to the part of the wave function inside the black hole, and the information that it contains about what had fallen into the black hole.

The first guess might be that this part of the wave function and the information it carries would emerge when the black hole finally disappears. However, information can't be carried for free, as one realizes when he gets the telephone bill. Information requires energy to carry it, and there's very little energy left in the final stages of a black hole. The only plausible way the information inside could get out would be if it emerged continuously with the radiation rather than waiting for this final stage. However, according to the picture of one member of the virtual particle pair falling in, and the other member escaping, one wouldn't expect the escaping particle to be related to what fell in, or to carry away information about it.

So the only answer would be that the information in the part of the wave function inside the black hole, gets lost. Such loss of information would have important implications for determinism. To start with, we've noted that even if we knew the wave function before the black hole disappeared, you can not run the Schroedinger equation backward and calculate what the wave function was before the black hole formed. What there was would depend in part on the bit of the wave function that got lost in the black hole. We're used to thinking we can know the past exactly; however, if information gets lost in black holes, this is not the case; anything could have happened.

In general, however, people such as astrologers and those who consult them, are more interested in predicting the future, than in retro…cting the past. At first glance in might seem that the loss of part of the wave function down the black hole would not prevent us from predicting the wave function outside the black hole. But it turns out that this loss does interfere with such a prediction, as we can see when we consider a thought experiment proposed by Einstein, Boris Predovski and Nathan Rosen, in the 1930's. Imagine that a radioactive atom decays and sends out two particles in opposite directions and with opposite spins.

An observer who looks only at one particle can't predict whether it would be spinning to the right or to the left, but if the observer measures it to be spinning to the right, then he can predict with certainty that the other particle would be spinning to the left, and vice-versa. Einstein thought that this proved that quantum theory was ridiculous; the other particle might be at the other side of the galaxy by now, yet one would instantaneously know which way it was spinning; however, most other scientists agreed that it was Einstein who was confused, not quantum theory.

The E-P-R experiment doesn't show that one is able to send information faster than light, that would be the ridiculous part. One can't choose that one's own particle would be measured to be spinning to the right, so one can't prescribe that the distant observer's particle should be spinning to the left. In fact, this thought experiment is exactly what happens with black hole radiation. The virtual particle pair will have a wave function that predicts that the two members would definitely have opposite spins. What we would like to do is to predict the spin and wave function of the outgoing particle, which we could do if we could observe the particle that's fallen in; but that particle is now inside the black hole, where its spin and wave function can't be measured. Because of this, it's not possible to predict the spin or wave function of the particle that escapes. It could have different spins and different wave functions with various probabilities, but it doesn't have a unique spin or wave function.

Thus, it would seem that our power to predict the future would be further reduced; the classical idea of LaPlace, that one could predict both the position and velocity of particles had to be modified when the uncertainty principle showed that one couldn't accurately measure both position and velocity. However, one could still measure the wave function and use the Schroedinger equation to predict what it should be in the future. This would allow one to predict with certainty one combination of position and velocity, which is half of what one could predict, according to LaPlace's ideas. We could predict with certainty that particles have opposite spins, but if one particle falls into the black hole, there is no prediction we can make with certainty about the remaining particle. This means that there isn't any measurement outside the black hole that can be predicted with certainty. Our ability to make definite predictions would be reduced to zero. So, maybe astrology is no worse at predicting the future than the laws of science.

Many physicists didn't like this reduction in determinism, and therefore suggested that information about what is inside could somehow get out of the black hole. For years it was just a piar's (?) hope that some way to save the information would be found. But in 1996, Andrew Strominger and Cumrun Vafa made an important advance. They chose to regard a black hole as being made up of building blocks called P-branes. I recall that one way of thinking about P-branes is as sheets that move through the three dimensions of space, and also through seven extra dimensions that we don't notice. In certain cases one could show that the number of waves on the P-brane is the same as the amount of information one would expect a black hole to contain. If particles hit the P-brane, they excite extra waves on the branes. Similarly, if waves moving in different directions on the P-branes come together at some point, they can create a peak so great that a bit of the P-brane breaks away, and goes off as a particle; thus, the P-branes can absorb and emit particles, like black holes.

One can regard the P-branes as an effective theory, that is, while we don't need to believe that there are actually little sheets moving through a flat space-time, black holes behave as if they were made up of such sheets; it's like water, which is made up of billions and billions of H_2O molecules with complicated interactions. But a smooth fluid is a very good, effective model, at least for certain calsses of black holes. For these classes, the P-brane model predicts exactly the same rate of emission that the virtual particle pair model predicts. However, there is a very important difference: in the P-brane model, information about what falls into the black hole would be stored in the wave function for the waves in the P-branes. The P-branes are regarded as sheets in flat space-time, and for that reason, time would flow forward smoothly, the paths of light rays won't be bent, and the information on the waves won't be lost. Instead, the information would eventually emerge from the black hole in the radiation from the P-branes. Thus, according to the P-branes model, we can use the Schroedinger equation to calculate what the wave function would be at later times. Nothing will get lost, and time will roll smoothly on. We will have complete determinism in a quantum sense.

So, which of these pictures is correct? Does part of the wave function get lost down the black hole, or does all the information get out again, as the P-brane model suggests? This is one of the outstanding questions in the theoretical physics today. Many people believe that recent work shows that information is not all lost; the world is safe and predictable, and nothing unexpected will happen. But it's not clear; if one takes Einstein's general theory of relativity seriously, one must allow the possibility that space-time ties itself in a knot and information gets lost in the folds.

Chapter 5

Protecting the Past

5.1 ***Is time travel possible? Could an advanced civilisation go back and change the past?***

My friend and colleague, Kip Thorne, with whom I've had a number of bets, is not one to follow the accepted line in physics just because every one else does. This has led him to have the courage to be the first serious scientist to discuss time travel as a practical possibility.

It is tricky to speculate openly about time travel; one risks either an outcry of the waste of public money being spent on something so ridiculous, or it demands that the research be classified for military purposes. After all, how could we protect ourselves against someone with a time machine? They might change history and rule the world…

There are only very few of us foolhardy enough to work on a subject that is so politically incorrect in physics circles. We disguise the fact by using technical terms that are code for time travel. The basis for all modern discussions of time travel is Einstein's general theory of relativity.

As we've seen in earlier chapters, the Einstein's equations made space and time dynamic, by describing how they were curved and distorted by the matter and energy in the universe. In general relativity, someone's personal time - as measured by the wrist watch - would always increase, just as it did in Newtonian theory, or the flat space-time of special relativity. But there was now the possibility that space-time could be warped so much, that you could go off in a space ship and come back before you set out. One way this could happen is if there were wormholes, tubes of space-time - mentioned in chapter 4, that connect different regions of space and time. The idea is that you could steer your space ship into one mouth of the wormhole and come out at the other mouth, in a different place and at a different time.

Wormholes (if they exist) would be the solution to the speed limit problem in space. It would take tens of thousands of years to cross the galaxy in a space ship that travelled at less than the speed of light, as relativity demands; but you might go through a wormhole to the other side of the galaxy and be back in time for dinner. However, one can show that, if wormholes exist, you could also use them to get back before you set out. So, you might think that, one could do something like blowing up his rocket on the launch pad to prevent your setting out in the first place.

This is a variation of the grandfather paradox. What happens if you kill you grandfather before your father was conceived? Of course, that is a paradox only if you have free will to do what you like when you go back in time. This book will not go into a philosophical discussion of free will; instead it will concentrate on whether the laws of physics allow space-time to be so warped that a macroscopic body such as a space ship can return to its own past. According to Einstein's theory, a space ship necessarily travels at less than the local speed of light, and follows what's called a time-light path through space-time. Thus one can formulate a question in technical terms: "does space-time admit time-light curves that are closed?" - that is they're returning to their starting point again and again. I shall refer to such paths as time loops.

There are three levels that we can try to answer to this question. The first is Einstein's general theory of relativity, which assumes that the universe has a well-defined history, without any uncertainty; for this classical theory we have a fairly complete picture. However, as we've seen, this theory won't be quite right, because we observe that matter is subject to uncertainty and quantum fluctuations. We can therefore ask the question about time travel on a second level, that of semi-classical theory; in this we consider matter to behave according to quantum theory, with uncertainty and quantum fluctuations, but space-time to be well defined in classical. Here the picture is less complete, but at least we have some idea of how to proceed.

Finally, there is the full quantum theory of gravity - whatever that may be. In this theory, where not only matter but also space and time themselves are uncertain and fluctuate, is not even clear how to pose the question of whether time travel is possible. Maybe, the best we can do is to ask how people in regions where space-time is nearly classical and free from uncertainty would interpret their measurements. Would they think that time travel had taken place in regions of strong gravity and large quantum fluctuations?

To start with the classical theory, the flat space-time of special relativity - relativity without gravity - doesn't allow time travel, nor that there are curved space-times that were known early on. It was therefore a great shock to Einstein, when in 1949, Kurt Gödel discovered a space-time that was a universe full of rotating matter, with time loops through every point.

5.2. *Gödel 's incompleteness theorem*

In 1931, the mathematician Kurt Gödel proved his incompleteness theorem about the nature of mathematics. The theorem states that within any formal system of axioms, such as present day's mathematics questions always persist that they can neither be proved or disproved on the basis of the axioms that define this system. In other words, Gödel showed that there are problems that cannot be solved by any set of rules or procedures. Gödel 's theorem set fundamental limits on mathematics. It came as a great shock to the scientific community, since it overthrown the widespread belief that mathematics was a coherent and complete system based on a single logical foundation.

Gödel 's theorem, together with Heisenberg's uncertainty principle and chaos theory, form a corset of limitations to scientific knowledge, that came to be appreciated only during the XXth century.

The Gödel solution required a chronological constant, which may or may not exist in nature. But other solutions were subsequently found without a cosmological constant. A particularly interesting case is one in which two cosmic strings move at high speed past each other.

5.3. *Cosmic strings*

Cosmic strings are long heavy objects with a tiny cross section, that may have been produced during the early stages of the universe. Once cosmic strings formed, they were further stretched by the expansion of the universe, and by now, a single cosmic string could cross over the entire length of our observable universe. The occurrence of cosmic strings is suggested by modern theories of particles, that predict that in the early hot stages of the universe, matter was in a symmetric phase, much like liquid water (which is symmetrical - the same at every point, in every direction), rather than like ice crystals, which have a discrete structure. When the universe cooled, the symmetry of the early phase could have been broken in many ways in distant regions; consequently, the cosmic matter would have settled into different ground states in those regions. Cosmic strings are the configurations of matter at the boundaries between these regions; their formation was therefore an inevitable consequence of the fact that different regions could not agree on their ground states.

Cosmic strings should not be confused with the strings from the strings theory, although they're not entirely unrelated. They're objects with length, but a tiny cross section. Their occurrence is predicted in some theories of elementary particles. The space-time outside a single cosmic string is flat. However, it's flat space-time, with a wedge cut out, with the sharp end at the string; it's like a cone. Take a large circle of paper and cut out a segment like a slice of pie (a wedge with its corner at the center of the circle); then discard the piece you've cut out and glue the cut edges of the remaining piece together, so that you have a cone. This represents the space-time in which the cosmic string exists. Notice that because the surface of the cone is the same flat sheet of paper with which you started, minus the wedge, you could still call it flat, except at the apex. You can recognize that there is curvature at the apex by the fact that the circle around the apex has a smaller circumference than a circle drawn at the same distance around the center of the original round sheet of paper. In other words, a circle around the apex is shorter than one would expect for a circle of that radius in flat space, because of the missing segment.

Similarly, in the case of a cosmic string, the wedge that is removed from flat space-time shortens circles around the string, but doesn't affect time or distances along the string. This means that space-time around a single cosmic string doesn't contain any time loops, so it's not possible to travel into the past. However, if there is a second cosmic string that is moving relative to the first, its time direction would be a combination of the time and space direction of the first. This means that the wedge that is cut out for the second string would shorten both distances in space and time intervals as seen by someone moving with the first string. If the cosmic strings are moving nearly at the speed of light relative to each other, the saving of time going around both strings can be so great that one arrives back before one set out. In other words, there are time loops that one could follow to travel into the past. The cosmic string space-time contains matter that has positive energy density, and is consistent with the physics we know; however, the warping that produces the time loops extends all the way out into infinity in space, and back to the infinite past in time. Thus, these space-times were created with time travel in them. We have no reason to believe that our own universe was created in such a warped fashion, and we have no reliable evidence of visitors coming from the future, un-discounting the conspiracy theory that UFOs are from the future, and that the government know and that it covers it up; its record of cover-ups is not that good.

I should therefore assume that there were no time loops in the distant past, or more precisely, in the past of some surface through space-time that I should call S. The question is: "could some advanced civilisation build a time machine ?" That is, could it modify the space-time to the future of S so that time loops appeared in a finite region? I say a finite region because no matter how advanced a civilisation becomes, it could presumably control only a finite part of the universe.

In science, finding the right formulation of a problem is often a key to solving it, and this was a good example. To define what was meant by a time machine, I went back to some early work of mine. Time travel is possible in a region of space-time in which there are time-light paths that move at less than the speed of light, but which, nevertheless, manage to come back to the place and time they started, because of the warping of space-time.

Since I've assumed that there were no time loops in the distant past, there must be what I called a time travel horizon, the boundary separating the region of time loops from the region without them. Time travel horizons are rather like black hole horizons; while the black hole horizon is formed by light ray that just miss falling into the black hole, a time travel horizon is formed by light rays on the verge of meeting up with themselves. I don't take as my criterion for a time machine, what I call a finitely generated horizon; that is a horizon which is formed by light rays that emerge from a bounded region; in other words, they don't come in from infinity or from a singularity, but originate form a finite region containing time loops - the sort of region our advanced civilisation is supposed to create.

In adopting this definition as the footprint for a time machine we have the advantage of being able to use the machinery that Roger Penrose and I developed to study singularities and black holes. Even without using Einstein's equation I can show that in general, a finitely generated horizon will contain a light ray that actually meets up with itself; that is a light ray that keeps coming back to the same point over and over again. Each time the light ray came around it will be more and more blue-shifted, so that the images get bluer and bluer; the wave crests of a pulse of light will get closer together, and the light will get around in shorter and shorter intervals of its time. In fact, a particle of light will have only a finite history, as defined by its own measure of time, even though it went round and round in a finite region and didn't hit a curvature, a singularity. One might not care if a particle of light completed its history in a finite time, but I can also prove that there would be paths moving at less than the speed of light that had only finite duration. These could be the histories of observers who would be trapped in a finite region before the horizon, and would go around and around, faster and faster, until they reach the speed of light in a finite time.

So, if a beautiful alien in a flying saucer invites you in her time machine, step with care; you might fall into one of these trapped repeating histories of only finite duration.

These results don't depend on the Einstein's equations, but only on the way space-time would have to warp to produce time loops in a finite region. However, we can now ask what kind of matter an advanced civilisation would have to use to warp space-time so as to build a finite sized time machine. Can it have positive energy density everywhere, as in a cosmic string space-time I described earlier? Cosmic string space-time didn't satisfy my requirement that the time loops appear in a finite region. However, one might think that this was just because the cosmic strings were finitely long. One might imagine that one could build a finite time machine using finite loops of cosmic strings, and have the energy density positive everywhere. It's a pity to disappoint people such as Kip who w anted to return to the past, but it can't be done with positive energy density everywhere. I can prove that to build a finite time machine, you need negative energy. Energy density is always positive in classical theory, so time machines of finite size are ruled out of this level. However, the situation is different in the semi-classical theory, in which one considers matter to behave according to quantum theory, but space-time to be well defined in the classical.

As we've seen, the uncertainty principle of the quantum theory means that fields are always fluctuating up and down, even in apparently empty space, and have an energy density that is infinite.

Thus, one has to subtract an infinite quantity to get the finite energy density that we observe in the universe. This subtraction can leave the energy density negative, at least locally. Even in flat space-time, one could find quantum states in which the energy density is negative locally, although the total energy is positive. One might wonder whether these negative values actually cause space-time to warp in the appropriate way to build a finite time machine, but it seems they must.

As we saw in chapter 4, quantum fluctuations mean that even apparently empty space is full of pairs of virtual particles that appear together, move apart then come back together and annihilate each other. One member of the virtual particle pair would have positive energy, and the other one negative energy. When a black hole is present, the negative energy member can fall in, and the positive energy member can escape to infinity, where it appears as radiation that carries positive energy away from the black hole. The negative energy particles falling into the black hole, cause it to lose mass and to evaporate slowly, with its horizon shrinking in size.

Ordinary matter with positive energy density has an attractive gravitational effect and warps space-time to bend light rays toward each other, just as the ball on a rubber sheet in chapter 2 always makes the smaller ball bearin gs curve toward it, never away. This would imply that the area of the horizon of a black hole could only increase with time, never shrink. For the horizon oh a black hole to shrink in size, the energy density on the horizon must be negative and warp space-time to make light rays diverge from each other.

This was something I first realised when I was getting into bed, soon after the birth of my daughter; I won't say how long ago that was, but I now have a grand son.

The evaporation of black holes shows that on a quantum level the energy density can sometimes be negative and warp space-time in the direction that will be needed to build a time machine. Thus we might imagine that some very advanced civilisation could arrange things so that the energy density is sufficiently negative to form a time machine that could be used by macroscopic objects such as space ships.

However, there is an important difference between a black hole horizon (0which is formed by light rays that just keep going) and the horizon in a time machine (which contains closed light rays that keep going around and around).

A virtual particle moving on such a closed path would bring its ground state energy back to the same point again and again. One would therefore expect the energy density to be infinite on the horizon - the boundary of the time machine, the region in which one could travel into the past. This is born out by explicit calculations, in a few backgrounds that are simple enough for exact calculations. It would mean that a per son or a space probe that tried to cross the horizon to get into the time machine, would get wiped out by a bolt of radiation.

So the future looks black for time travel, or should I say blindingly white.

The energy density of matter depends on the state it is in. so it is possible that an advanced civilisation might be able to make the energy density finite on the boundary of the time machine by freezing out, or removing the virtual particles that go around and around in a closed loop. It's not clear how ever, that such a time machine would be stable. The least disturbance such as someone crossing their horizon to enter the time machine might set off circulating virtual particles and trigger a bolt of lightning. This is a question that physicists should be free to discuss without being laughed to scorn.

Even if it turns out that time travel is impossible, it's important that we know why it is impossible. To answer that question definitively, we need to consider quantum fluctuations; not only of matter fields, but of space-time itself.

One might expect that these would cause a certain fuzziness in paths of light rays and in the whole concept of time ordering. In deed, one can regard the radiation from black holes as leaking out because quantum fluctuations of space-time mean that the horizon is not exactly defined. Because we don't yet have a complete theory of quantum gravity it's difficult to say what the effects of quantum fluctuations should be. Nevertheless, we can hope to get some pointers from the Feynmann's sum over histories, described in chapter 3.

Each history would be a curved space-time with matter fields in it. Since we're supposed to sum over all possible histories, not just those that satisfy some equations, the sum must include space-times that are warped enough for travel into the past.

So, the question is "why isn't time travel happening everywhere?"

The answer is that time travel is in deed taking place on a microscopic scale, but we don't notice it. If someone applies the Feynmann's sum over histories idea to a particle, one has to include histories in which a particle travels faster than light and even backward in time. In particular, there will be histories in which the particle goes around and around on a closed loop in time and space. It will be like the film "groundhounds day" in which a reporter has to live the same day over and over again.

One can't observe such particles with closed loops histories directly with a particle detector; however, their indirect effects have been measured in a number of experiments. One is a small shift in the light given off by H2 atoms, caused by electrons moving on closed loops. Another is a small force between parallel metal plates caused by the fact that there are slightly fewer closed loops histories that can fit between the plates compared to the region outside; another equivalent interpretation of the Casimir effect. Thus, the existence of closed lops histories is confirmed by experiment.

One might dispute whether close loop particle histories has anything to do with the warping of space-time, because they occur even in fixed background such as flat space. But in recent years we found that phenomena in physics often have dual, equally valid descriptions.

One can equally well say that a particle moves on a closed loop on a fixed background, or that the particle stays fixed and the space-time fluctuates around it. It's just a question of whether you do the sum over particle paths first and then the sum over curved space-times, or vice-versa.

It seems therefore that quantum theory allows time travel on a microscopic scale; however, this is not much for SF purposes such as going back and killing your grand father.

The question therefore is "can the probability in the sum over histories peaked/picked around space-times with macroscopic time loops"? One can investigate this question by studying the sum over history of matter fields in a series of background space-times that get closer and closer to admitting time loops. One would expect something dramatic to happen when time loops first appear, and this is born out in the simple example that I studied with my student Michael Cassidy. The background space-times in the series we've studied were closely related to what is called the Einstein's universe - a space-time that Einstein proposed when he believed that the universe was static and unchanging in time (neither expanding, nor contracting - see chapter 1).

In the Einstein's universe, time runs from the infinite past to the infinite future. The space directions however are finite and close on themselves, like the surface of the Earth, but with one more dimension.

One can picture this space-time as a cylinder with the long axis being the time direction and the cross section being the three space directions.

The Einstein's Universe doesn't represent the universe we live in because it's not expanding; nevertheless, it is a convenient background to use when discussing time travel, because it's simple enough that we can do the sum over histories.

Forgetting about time travel for a moment, consider matter in an Einstein universe that is rotating about some axis. If you are on the axis you can remain at the same point of space, just as you do when standing at the center of a children's carousel. But if you are not on the axis, you would be moving through space as you rotated about the axis; the farther you were from the axis, the faster you'd be moving.

So if the universe were infinite in space, points sufficiently far from the axis would have to be rotating faster than light. However, because the Einstein universe in finite in the space directions, there is a critical rate of rotation below which no particle in the universe is rotating faster than light.

Now, consider the sum over particle histories in a rotating Einstein universe. When the rotation is slow, there are many paths a particle can take using a given amount of energy. Thus, the sum over all particle histories in this background gives a large amplitude. This means that the probability of this background would be high in the sum over all curved space-time histories; that is, it's among the more probable histories. However, as the rate of rotation of the Einstein universe approaches the critical value, so that its outer edges are moving at speeds approaching the speed of light, there is only one particle path that is classically allowed on that edge, namely, one that is moving at the speed of light. This means that the sum over particle histories would be small; thus, the probability of these backgrounds would be low on the sum over all curved space-time histories; that is, they are the least probable.

What do rotating Einstein universes have to do with time travel and time loops? The answer is that they are mathematically equivalent to other backgrounds that do admit time loops. These other backgrounds are universes that are expanding in two space directions; the universes are not expanding in the third space direction, which is periodic. That is to say, if you go a certain distance in this direction, you get back to where you started. However, each time you do a circuit of the third space direction, your speed in the first or second space direction gets boosted. If the boost is small, there are no time loops. However, consider a sequence of backgrounds with increasing boosts in speed. At a certain critical boost, time loops will appear. Not surprisingly, this critical boost corresponds to the critical rate of rotation of the Einstein universes.

Since the sum over histories calculations in these backgrounds are mathematically equivalent, one can conclude that the probability of these backgrounds goes to zero as they approach the warping needed for time loops. In other words, the probability of having sufficient warping for a time machine is zero.

This supports what I have called the chronology protection conjecture, mentioned at the end of chapter 2, that the laws of physics conspire to prevent time travel by macroscopic objects.

Although time loops are allowed by the sum over histories, the probabilities are extremely small. Based on the duality arguments I mentioned earlier, I estimate the probability that Kip Thorne could go back and kill his grand father is less than 1 in 10^60.

As gambling men, Kip and I will bet on odds like that; the trouble is, we can't bet each other, because we're now on the same side. On the other hand, I wouldn't take a bet with anyone lese; he might be from the future and know that time travel worked.

You might wonder if this chapter is part of a government cover-up on time travel; you might be right...

Chapter 6

Our Future

6.1 ***"Star Trek" or not.***
How biological and electronic life will go on developing in complexity at an ever-increasing rate

The reason Star Trek is so popular is because it's a safe and comforting vision of the future. I'm a bit of a Star Trek fan myself, so I was easily persuaded to take part in an episode in which I played poker with Newton, Einstein and Cmdr. Data.

Now I beat them all but unfortunately there was a red alert so I never collected my winnings.

Star Trek shows a society that far in advance of ours in science, in technology and in political organisation. The last might not be difficult.

There must have been great changes, whether accompanying tensions and upsets in the time between now and then; but in the period we're shown, science, technology and the organisation of society are supposed to have achieved a level of near perfection.

I want to question this picture and ask if we will ever reach a final steady state in science and technology.

At no time, in the 10000 years or so since the last ice age, has the human race been in the state of constant knowledge and fixed technology. There have been a few setbacks like the dark ages, after the fall of the Roman Empire.

But the world's population, which is a measure of our technological ability to preserve life and feed ourselves has risen steadily, with only a few hick-ups such as the black death.

In the last 200 years population growth has became exponential, that is the population grows by the same percentage each year. Currently the rate is about 1.9 % a year; that may not sound like very much but it means that the world's population doubles every 40 years.

Other measures of technological development in recent times are electricity consumption and the number of scientific articles. They too show exponential growth, with doubling times of less than 40 years. There's no sign that scientific or technological development will slow down and stop in the near future; certainly not by the time of Star Trek, which is supposed to be not that far in the future.

But if the population growth and the increase in consumption in electricity continue at their current rates, by 2600 the world's population will be standing shoulder to shoulder and electricity use will make the earth glow red hot.

If you stacked all the new books being published next to each other you will have to move at 90 mph just to keep up with the end of the line. Of course, by 2600, new artistic and scientific work will come in electronic forms rather than as physical books and papers.

Nevertheless, if the exponential growth continued, there will be 10 papers a second in my kind of theoretical physics and no time to read them.

Clearly, the present exponential growth can't continue indefinitely; so what will happen?

One possibility is that we'll wipe ourselves out completely by some disaster, such as a nuclear war. There's a sick joke that the reason why we've not been contacted by extraterrestrials, is that when a civilisation reaches our state of development, it becomes unstable and destroys itself.

However, I'm an optimist; I don't believe that the human race has come so far just to snuff itself out when things are getting interesting.

The Star Trek vision of the future that we've achieved an advanced but essentially static level may come true in respect of our knowledge of the basic laws that the universe.

As I shall describe in the next chapter, there may be an ultimate theory that we'll discover in a not too distant future. This ultimate theory, if it exists, will determine whether the Star Trek dream of warp drive can be realised.

According to present ideas, we should have to explore the galaxy in a slow and tedious(?) manner, using space ships, travelling slower than light.

But since we don't yet have a complete unified theory, we can't quite rule out warp drive. On the other hand, we already know the laws that hold in all but the most extreme situations. The laws that govern the crew on the Enterprise, if not the space ship itself.

Yet it doesn't seem that we'll ever reach a steady state in the uses we make of these laws, or in the complexity of the systems that we can produce with them. It is with this complexity that the rest of this chapter will be concerned.

By far the most complex systems that we have are our own bodies. Life seems to have originated in the primordial oceans that covered the earth $4*10^9$ years ago. How this happened, we don't know. It may be that random collisions between atoms built up macromolecules that could reproduce themselves and assemble themselves into more complicated structures.

What we do know is that by $3.5*10^9$ years ago, the highly complicated DNA molecule had emerged. DNA is the basis for all life on Earth. It has a double helix structure like a spiral staircase, which was discovered by Frances Crek (?) and James Watson in the Cavendish Lab at Cambridge in 1953.

The two strands of the double helix are linked by pairs of nucleic acids, like the threads in the spiral staircase. There are four kinds of nucleic acids: adenosine, cytosine, guanine, tyrosine. The order in which different nucleic acids occur along the spiral staircase carries the genetic information that enables the DNA molecule to assemble in organism around it and reproduce itself.

As the DNA makes copies of itself, there are occasional errors in the order of the nucleic acids along the spiral. In most cases, the mistakes in copying make the DNA either unable or less likely to reproduce itself; meaning that such genetic errors or mutations, as they're called, will die out. But in a few cases, the error, or mutation, will increase the chances of the DNA surviving and reproducing.

Such changing in the genetic code will be favourite. This is how the information contained in the sequence of nucleic acids gradually evolves and increases in complexity. Because biological evolution is basically a random walk in the space of all genetic possibilities, it's been very slow. The complexity, or number of bits of information that is coded in the DNA, is roughly the number of nucleic acids in the molecule.

For the first 2 billion years or so, the rate of increase in complexity must have been of the order of one bit of information every hundred years. The rate of increase in DNA complexity gradually rose to about one bit a year over the last few million years. But then, about six or eight thousand years ago, a major new development occurred. We developed written language.

This meant that information could be passed from one generation to the next without having to wait for the very slow process of random mutations and natural selection to code it into the DNA sequence. The amount of complexity increased enormously. A simple paperback romance could hold as much information as the difference in the DNA between apes and humans; and a 30 volume encyclopaedia could describe the entire sequence of human DNA.

Even more important, the information in books can be updated rapidly; the current rate at which human DNA is being updated by biological evolution is about 1 bit a year; but there are $2*10^5$ new books published each year; a new information rate of over 10^6 bits a second. Of course, most of this information is garbage, but even if 1 bit in a million is useful, that is still a hundred thousand times faster than biological evolution.

This transmission of data through external, non-biological means, has led the human race to dominate the world and to have an exponentially increasing population. But now we're at the beginning of a new era in which we'll be able to increase the complexity of our internal record, the DNA, without having to wait for the slow process of biological evolution.

There has been no significant change in the human DNA in the last 10000 years, but it is likely that we'll be able to completely redesign it in the next thousand. Of course, many people will say that genetic engineering of humans should be banned, but it's doubtful we'll be able to prevent it. Genetic engineering of plants and animals will be allowed for economic reasons, and someone is bound to try it on humans.

Unless we have a totalitarian world order, someone, somewhere, will design improved humans. Clearly, creating improved humans will create great political and social problems with respect to unimproved humans.

My intention is not to defend human genetic engineering, as a desirable development, but just to say it's likely to happen, whether we want it or not.

This is the reason why I don't believe SF like Star Trek where people 400 years into the future are essentially the same as we are today.

I thing the human race and it's DNA will increase its complexity quite rapidly. We should recognize that this is likely to happen, and consider how we'll deal with it. In a way, the human race needs to improve its mental and physical qualities if it's to deal with the increasingly complex world around it meet new challenges such as the space travel. Humans also need to increase their complexity if biological systems are to keep ahead of electronic ones.

At the moment computers have the advantage of speed but they show no sign of intelligence. This is not surprising, because our present computers are less complex that the brain of an earthworm - a species not noted for its intellectual powers. The computers obey what is known as More's law: their speed and complexity doubles every 18 months; it's one of those exponential growths that clearly can't continue indefinitely.

However, it will probably continue until computers will have the complexity similar to that of the human brain. Some people say that computers can never show true intelligence, whatever that may be; but it seems to me that if very complicated chemical molecules can operate in humans, then equally complex electronic circuits can also make computers act in an intelligent way. And if they're intelligent, they can presumably design computers that have an even greater complexity.

Will this increase in biological and electronically complexity go on forever, or is there a natural limit? On the biological side, the limit of human intelligence up to now is been set by the size of the brain that will pass through the birth canal. Having watched my three children being born, I know how difficult it is for the head to get out.

But within the next hundred years I expect we'll be able to grow babies outside the human body; so this limitation will be removed.

Ultimately, however, increases in the size of the human brain through genetic engineering will come up against the problem that the body's chemical messengers, responsible for our mental activity, are relatively slow moving. This means that further increases in the complexity of the brain, will be at the expense of speed. We can be quick-witted(?), or very intelligent, but not both. Still, I think we can become more intelligent than most of the people in Star Trek, not that that might be difficult.

Electronic circuits have the same complexity vs. speed problem as the human brain; in this case, however, the signals are electrical, not chemical, and travel at the speed of light, which is much higher.

Nevertheless, the speed of light is already a practical limit on the design of faster computers. One can improve the situation, by making the circuits smaller, but ultimately there will be a limit set by the atomic nature of matter.

Still, we have some way to go before we meet that barrier.

Another way in which electronic circuits can increase their complexity while maintaining speed, is to copy the human brain; the brain does not have a single CPU, that processes each command in sequence. Rather, it has millions of processors working together at the same time. Such massively parallel processing will be the future for electronic intelligence as well.

Assuming we don't destroy ourselves in the next hundred years, it is likely that we'll spread out, first to the planets in the solar system, and then to the nearby stars. But it won't be like Star Trek or Babylon 5, with a new race of nearly human beings in almost every stellar system. The human race has been in its present form for only 2 million years out of 15 billion years or so since the big-bang.

6.2 *The biological-electronic interface*

Within two decades, a 1000 $ computer may be as complex as the human brain. Parallel processors could mimic the way our brain works, and make computers act in intelligent and conscious ways. Neural implants may allow a much faster interface between the brain and computers, disolving the distance between biological and electronic intelligence.

In the near future, most business transactions will probably be made between cyber-personalities via the www. Within a decade, many of us may choose to live a virtual existence on the net, forming cyber-friendships and relationships.

Our understanding of the human genome will undoubtably create great medical advances; but they will also enable us to increase the complexity of the human DNA structure significantly. In the next few hundred years, human genetic engineering will replace biological evolution, redesigning the human race, and posing entirely new ethical questions.

Space travel beyond our solar system will probably require either genetically engineered humans, or unmanned computer controlled probes.

So even if life develops in other solar systems, the chances of catching it at a recognizably human stage, are very small. Any alien life we encounter will either be much more primitive, or much more advanced.

If it is more advanced, why hasn't it spread over the galaxy, and visited earth. If aliens had come here, it should have been more obvious; more like the film "Independence Day", than "ET".

So how does one account for the lack of extraterrestrial visitors? It could be that there is an advanced race out there, which is aware of our existence, but is leaving us to stew in out own primitive juices.

However, it is doubtful that it will be so considerate to allow a life form. Do most of us worry about how many insects and earthworms we squash under foot.

A very reasonable explanation is that there is a very low probability, either of life developing on other planets, or of that life developing intelligence.

Because we claim to be intelligent, although perhaps without much ground, we don't see intelligence as an inevitable consequence of evolution. However, one can question that. It's not clear that intelligence has much survival value; bacteria do very well without intelligence, and will survive us, if our so-called intelligence will cause us to wipe ourselves out in a nuclear war.

So, as we explore the galaxy, we may find primitive life, but we're not likely to find beings like us.

The future of science won't be like the comforting picture painted in Star Trek - a universe populated by many humanoid races with an advanced but essentially static science and technology; instead I think we'll be on our own, but rapidly developing on biological and electronic complexity.

Not much of this will happen in the next hundred years, which is all we can reliably predict; but by the end of the next millennium, if we get there, the difference from Star Trek will be fundamental.

Chapter 7

Brane new World

7.1 ***Do we live on a brane, or are we just holograms? How will our journey of discovery proceed in the future?***

Will we succeed in our quest for a complete unified theory that will govern the universe and everything that it contains? In fact, as described in chapter 2, we may have already identified the theory of everything (TOE) as M-Theory.

This theory doesn't have a single formulation, at least as far as we know, instead we've discovered a network of apparently different theories, which all seem to be approximations to the same underlying fundamental theory, in different limits; just as Newton's theory of gravity is an approximation to Einstein's general theory of relativity, in the limit that the gravitational field is week.

M-Theory is like a jigsaw. It's easy to identify and fit together the pieces around the edges of the jigsaw - the limits of M-theory where some quantity ... or either is small (?)

We now have a fairly good idea of these edges, but there is still a gaping hole at the centre of the M-theory, where we don't know what's going on. We can't really claim to have found the theory of everything until we have filled that hole.

What is in the centre of the M-theory? Will we discover dragons or something equally strange like on old maps of unexplored lands? Our experience in the past suggests we're likely to find unexpected new phenomena whenever we extend the range of our observations to smaller scale.

At the beginning of the 20-th century we understood the workings of nature on the scales of classical physics, which is good for interstellar distances down to about 1/100 of a millimetre.

Classical physics assumes that matter is a continuous medium, with properties like elasticity and viscosity; but evidence began to emerge that matter is not smooth but grainy; it's made of tiny building blocks called atoms.

The word atom comes from Greek and means indivisible; but it was soon found that atoms consisted of electrons orbiting a nucleus made up of protons and neutrons.

The work on atomic physics, in the first 30 years of the 20-th century, took our understanding down to lengths of a millionth of a millimetre.

Then we discovered that protons and neutrons were made of even smaller particles called quarks.

Our recent research on nuclear and high energy physics has taken us to length scales that are smaller by a further factor of a billion.

It might seem that we can go on forever discovering structures on smaller and smaller length scales.

However, there is a limit to this series, as there is to the series of Russians dolls within Russian dolls; eventually, one gets down to a smallest doll which can't be taken apart any more.

In physics, the smallest doll is called the Planck length. To probe to shorter distances would require particles of such high energy that they were inside black holes.

We don't know exactly what the fundamental Planck length is in M-theory but it might be as small as a millimetre divided by 100000 billion billion billion.

We are not about to build particle accelerators that can probe to distances that small. They would have to be larger than the Solar System, and they're not likely to be approved in the present financial climate.

However, there has been an exciting new development that means we might discover at least some of the dragons of M-theory more easily and cheaply.

As explained in chapters 2 and 3, in the M-theory network of mathematical models, spate-time has 10 or 11 dimensions. Up to recently, it was thought that the 6 or 7 extra dimensions would all be curled up very small; it would be like a human hair.

If you look at a hair under a magnifier glass, you can see it has thickness; but to the naked eye it just appears like line with a length but no other dimension.

Space-time may be similar. On human, atomic, or even nuclear physics length scales, it may appear 4-D, and nearly flat. On the other hand, if we probe to very short distances, using extremely high energy particles, we should see that space-time was 10 or 11-D.

7.2 *Brane worlds*

If only additional dimensions were very small, it would be very difficult to observe them. However, there has recently been a suggestion that one or more of the extra dimensions might be comparatively large or even infinite.

This idea has the great advantage, at least to a positivist like me, that it may be testable by the next generation of particle accelerators or by sensitive short range measurements of the gravitational force.

Such observations could either falsify the theory or experimentally confirm the presence of other dimensions.

Large extra dimensions are in exciting new development in our search of the ultimate model of theory. They would imply that we lived in a brane world, a 4-D surface, or brane, in a higher dimensional space-time.

Matter and non-gravitational forces, like the electric force, would be confined to the brane; thus, everything not involving gravity would behave as it would in four dimensions.

In particular, the electric force between the nucleus of an atom and the electrons orbiting around it, would fall off with distance at the right rate for atoms to be stable, against the electrons falling into the nucleus.

This would be in accordance with the anthropic principle, that the universe must be suitable for intelligent life; if atoms weren't stable, we wouldn't be here to observe the universe and ask "why it appears 4-D".

On the other hand, gravity, in the form of curved space would permeate the whole bulk of the higher dimensional space-time; this would mean that gravity would behave differently for other forces we experience.

Because gravity would spread out in the extra dimensions, it would fall off more rapidly with distance than one would expect. If this more rapid fall off of the gravitational force extended to astronomical distances, we would have noticed its effect on the orbits of the planets. In fact, they would be unstable, as was remarked in chapter 3. The planets would either fall into the Sun, or escape to the dark and cold of interstellar space.

However, this would not happen if the extra dimensions ended on a brane not that far away from the brane on which we live. Then, for distances greater than the separation of the branes, gravity would not be able to spread out freely, but would effectively be confined to the brane - like the electric forces - and fall off at the right rate for the planetary orbits.

On the other hand, at distances less than the separation of the branes, gravity would vary more rapidly. The very small gravitational field between heavy objects has been measured accurately in the lab; but the experiments so far would not have detected the effect of branes separated by less than a few inches.

New measurements are now being made at shorter distances.

In this brane-world, we would live on one brane, but there would be another shadow brane nearby. Because light would be confined to the branes, it would not propagate through the space between, we couldn't see the shadow world, but we would feel the gravitational influence of the matter on the shadow brane.

In our brane, such gravitational forces would appear to be produced by sources that were truly dark, and that the only way that we could detect them is through their gravity.

In fact, in order to explain the rate at which stars orbit the centre of our galaxy, it seems there must be more mass than is accounted for by the matter we observe.

7.3 *Evidence for dark matter*

Various cosmological observations strongly suggest that there should be much more matter in our galaxy, and other galaxies, than we see.

The most convincing of these observations is that stars at the outsquirts of the spiral galaxies, like our own Milky Way, orbit far too fast to be held on their orbits only by the gravitational attraction of all the stars we observe.

We've known, sine the 1970's, that there is a discrepancy between the observed rotational velocities of stars in the outer regions of the spiral galaxies, and the orbit velocities that one would expect according to Newton's laws, from the distribution of the visible stars in the galaxy.

This indicates that there should be much more matter in the outer parts of the spiral galaxies.

7.4 *The nature of dark matter*

Cosmologists now believe that while the central parts of spiral galaxies consist largely or ordinary stars, their outsquirts are dominated by dark matter that we can't see directly. One of the fundamental problems for cosmology today is to discover the dominant form of dark matter in the outer regions of galaxies.

Before the 1980's there was usually assumed that this dark matter was ordinary matter comprised of protons, neutrons and electrons, in some not readily detectable form - perhaps gas clouds or MACHOs (massive compact halo objects) like white dwarfs of neutron stars, or even black holes.

However, recent study on the formation of galaxies has led the cosmologists to believe that a significant fraction of the dark matter must be in a different form than for ordinary matter. Perhaps it arises from the masses of very light elementary particles such as axions and neutrinos.

It may even consist of more exotic species of particles such as WIMPs (weakly interacting massive particles) that are predicted by modern theories of elementary particles, but have not yet been detected experimentally. This missing mass might arise from some exotic species of particle in our world such as wimps, or axions (very light elementary particles).

But missing mass could also be evidence of the existence of a shadow world with matter in it. Maybe it contains shadow human beings wondering about the mass that seems to be missing from their world to account for the orbits of shadow stars around the centre of the shadow galaxy.

Instead of extra dimension ending on the second brane, another possibility is that they are infinite but highly curved, like a saddle.

Lisa Randall and Ramon Sundrum (?) showed that this kind of curvature would act like a second brane; the gravitational influence of an object on the brane would be confined to a small neighbourhood of the brane and not spread out to infinity in the extra dimensions.

Like in the shadow brane model, the gravitational field would have the right long distance fall off to explain planetary orbits and lab measurements of the gravitational force. But gravity would vary more rapidly at short distances.

There is, however, an important difference between this Randall-Sundrum model and the shadow brane model. Bodies that move under the influence of gravity will produce gravitational waves, ripples of curvature that travel through space-time at the speed of light. Like the electromagnetic waves of light, gravitational waves should carry energy - a prediction that has been confirmed by observations of the binary pulsar PSR ...

7.5 *Binary pulsars*

General relativity predicts that heavy bodies moving under the influence of gravity emit gravitational waves. Like light waves, gravitational waves carry energy away from the objects that emit them. However, the rate of energy loss is usually extremely low, hence very difficult to observe. For instance of gravitational waves is causing the Earth to slowly spiral in towards the Sun, but it would take 10^{27} years for them to collide.

But in 1975, Russel Holls (?) and Joseph Taylor discovered the binary pulsar PSR 1913+16, a system consisting of two compact neutron stars orbiting each other with a maximum separation of only one solar radius.

According to general relativity, the rapid motion means that the orbital period of the system should decrease on a much shorter time-scale, because of the emission of a strong gravitational wave signal.

The change predicted by general relativity is in excellent agreement with careful observation by Holls and Taylor of the orbital parameters, which indicate that since 1975, the period has shortened by more than 10 seconds.

In 1993 they were awarded the Nobel Prize for this confirmation of general relativity.

If we indeed live on a brane in a space-time with extra dimensions, gravitational waves generated by the motion of bodies on the brane would travel off into the other dimensions. They would be reflected back if there were a shadow brane and trap between the two branes.

On the other hand, if there was only a single brane, and the extra dimensions went on forever, like in the Randall-Sundrum model, gravitational waves could escape altogether and carry away energy from our brane world.

This would seem to breach one of the fundamental principles of physics - the law of conservation of energy; the total amount of energy remains the same. However, it appears to be a violation only because our view of what is happening is restricted to the brane.

An angel, who could see the extra dimensions, would know that the energy is the same, just more spread out.

The gravitational waves produced by two stars orbiting each other would have a wavelength which would be much longer than the radius of the saddle shape curvature in the extra dimensions.

This would mean they would tend to be confined to a small neighbourhood of the brane, like gravitational force, and wouldn't spread out much in the extra dimensions or carry away much energy from the brane.

On the other hand, gravitational waves that were shorter than the scale on which the extra dimensions are curved, would escape easily from the vicinity of the brane.

The only sources of significant amounts of short gravitational waves are likely to be black holes. A black hole on the brane would extend to a black hole in the extra dimensions. If the black hole is small, it would be almost round; that is, it would reach about as far into the extra dimensions as its size on the brane.

On the other hand, a large black hole on the brane, would extend to a black pancake, which is confined to a vicinity of the brane on which is much less thick in the extra dimensions than it is wide on the brane.

As explained in chapter 4, quantum theory means that black holes won't be completely black. They'll emit particles in radiation of all kinds like hot bodies. The particles in radiation like light would be emitted along the brane because matter and non-gravitational forces like electricity would be confined to the brane.

However, black holes also emit gravitational waves. These would not be confined to the brane but will travel into the extra dimensions as well. If the black hole was large and pancake-like, the gravitational waves would stay near the brane; this would mean that the black hole would loose energy, and therefore mass, by $E = mc^2$, at the rate one would expect for a black hole in 4-D space-time.

The black hole will therefore slowly evaporate and shrink in size until it became smaller than the radius of curvature of the saddle-like extra dimensions. At this point, the gravitational waves emitted by the black hole will begin to escape freely into the extra dimensions.

To someone on the brane, the black hole - or dark star, as Mitchel called it - would appear to be emitting dark radiation (radiation that can't be observed directly on the brane, but whose existence could be inferred by the fact that the black hole was loosing mass). It would mean that the final burst of radiation from an evaporating black hole would appear less powerful than it actually was.

This could be why we've not observed bursts of gamma rays that could be ascribed to dying black holes, though the other more prosaic explanation would be that there aren't many black holes with mass low enough to evaporate at the age of the universe so far.

The radiation from brane-world black holes arises from quantum fluctuations of particles on and off a brane. But branes, like everything else in the universe, will be subject to quantum fluctuations themselves. These can cause branes to appear and disappear spontaneously. The quantum creation of a brane would be a bit like the formation of bubbles of steam in boiling water.

Liquid water consists of billions and billions of H_2O molecules, packed together, with coupling between nearest neighbours. As the water is heated up, the molecules move faster and bounce off each other. Occasionally, these collisions will give molecules such high velocities that a group of them will break free of their bonds and form a little bubble of steam surrounded by water.

The bubble will then grow or shrink in a random manner, with more molecules from the liquid joining the steam or vice-versa. Most small bubbles of steam will collapse to liquid again, but a few would grow to a certain critical size, beyond which bubbles are almost certain to continue to grow. It is these large expanding bubbles that one observes when water boils.

The behaviour of brane-worlds would be similar. The uncertainty principle would allow brane-worlds to appear from nothing, as bubbles with the brane forming the surface of the bubble, and the interior being the higher dimensional space. Very small bubbles would tend to collapse again to nothing; but a bubble that grew by quantum fluctuations beyond a certain critical size, would be likely to keep on growing.

People, such as us, living on a brane - the surface of the bubble - would think the universe was expanding; it would be like painting galaxies on the surface of a balloon and blowing it up. The galaxies would move apart but no galaxy would be picked out as the centre of the expansion. Let's hope that there is no one with a cosmic pin to deflate the bubble.

According to the no boundary proposal, described in chapter 3, the spontaneous creation of a brane-world would have a history in imaginary time, which was like a nutshell - that is, it would be a 4-D sphere, like the surface of the Earth, but with 2 more dimensions; the important difference is that the nutshell described in chapter 3 was essentially hollow. The 4-D sphere wouldn't have been the boundary of anything, and the other 6 or 7 dimensions of the space-time that M-theory predicts, would all be curled up even smaller than the nutshell.

On the new brane-world picture, however, the nutshell will be filled; the history in imaginary time of the brane on which we live would be a 4-D sphere which would be the boundary of a 5-D bubble, with the remaining 5 or 6 dimensions curled up very small. This history of the brane in imaginary time would determine its history in real time; in real time the brane would expand in an accelerated inflationary manner, like that described in chapter 3. A perfectly smooth and round nutshell would be the most probable history of the bubble in imaginary time.

However, it will correspond to a brane that expanded forever in an inflationary way in real time. Galaxies wouldn't form on such a brane; and so intelligent life would not have developed. On the other hand imaginary time histories, that are not perfectly smooth and round would have somewhat lower probabilities, but could correspond to real time behaviour in which the brane had a phase of accelerating inflationary expansion at first, but then began to slow down.

During this decelerating expansion, galaxies could have formed and intelligent life might have developed. Thus, according to the anthropic principle described in chapter 3, it is only the slightly hairy nutshells which would be observed by intelligent beings asking "why the origin of the universe wasn't perfectly smooth".

As the brane expanded, the volume of the higher dimensional space inside would increase; eventually there would be an enormous bubble surrounded by the brane on which we live. But, do we really live on a brane? According to the idea of the holography, described in chapter 2, information about what happens in a region of space-time can be encoded on its boundary.

So, maybe we think we live on a 4-D world, because we are shadows cast on the brane by what is happening in the interior of the bubble.

However, from a positivist viewpoint, one can't ask "which is reality - brane or bubble". They are both mathematical models that describe the observations. One is free to use whichever model is most convenient.

What is outside the brane? There are several possibilities.

1 - There may be nothing outside; although a bubble of steam has water outside it, this is just an analogy to help us visualize the origin of the universe. One could imagine a mathematical model that was just a brane with a higher dimensional space inside but absolutely nothing outside, not even empty space. One can calculate what the mathematical model predicts without reference to what is outside.

2 - One could have a mathematical model on which the outside of the bubble was glued to the outside of a similar bubble. This model is actually mathematically equivalent to the possibility discussed above, that there is absolutely nothing outside of the bubble. But the difference is psychological. People feel

happier being placed in the centre of space-time rather than on its edge. But for a positivist, possibilities 1 and 2 are the same.

3 - the bubble might expand into a space that was not a mirror image of what was inside the bubble. This possibility is different to the two discussed above and is more like the case of boiling water. Other bubbles could form and expand. If they collided and merged with the bubble in which we lived, the results could be catastrophic. It is even been suggested that the Big-Bang itself may have been produced by a collision between branes.

Brane-world models like this are a hot topic of research; they are highly speculative, but they offer new kinds of behaviour which could be tested by observation. They could explain why gravity seems to be so weak. Gravity might be quite strong in the fundamental theory, but the spreading of the gravitational force in the extra dimensions would mean it would be weaker at large distances on the brane on which we live.

A consequence of this would be that the Planck length - the smallest distance to which we can probe without creating a black hole - would be quite a lot larger than it would appear from the weakness of gravity on our 4-D brane.

The smallest Russian doll wouldn't be so tiny after all, and might be in reach of particle accelerators of the future. In fact, we might have already discovered the smallest doll, the fundamental Planck length, if the US hadn't gone through a fit'n'fill in por (?) in 1994 and cancel the SSC (the superconducting supercollider), eventhough it was half-built.

Other particle accelerators, such as the LHC (Large Hadron Colider) in Geneva are now being built. With them and other observations, such as the CMBR, we may be able to determine whether or not we live on a brane. If we do, it will presumably be because the anthropic principle pricks out brane models from the vast zoo of universes allowed by M-theory.

We could well paraphrase the Miranda in Shakespeare as the Tempist: "Oh, brane new world, that has such creatures in it".

That is the Universe in a Nutshell.

CONTENT:

Chapter 1
A Brief History of Relativity
How Einstein laid the foundations of the two fundamental theories of the twentieth century.
General Relativity and Quantum Theory

Chapter 2
The Shape of Time
Einstein's General Relativity gives time a shape. How this can be reconciled with Quantum Theory

Chapter 3
The Universe in a Nutshell
The universe has multiple histories, each of which is determined by a tiny nut

Chapter 4
Predicting the Future
How the loss of information in the black holes may reduce our ability to predict the future

Chapter 5
Protecting the Past
Is time travel possible? Could an advanced civilization go back and change the past?

Chapter 6
Life in the Future-Star Trek or Not?
How biological and electronic life will go on developing in complexity at an ever increasing rate

Chapter 7
Brane New World
The way ahead-brane worlds, biography, and the theory of Everything