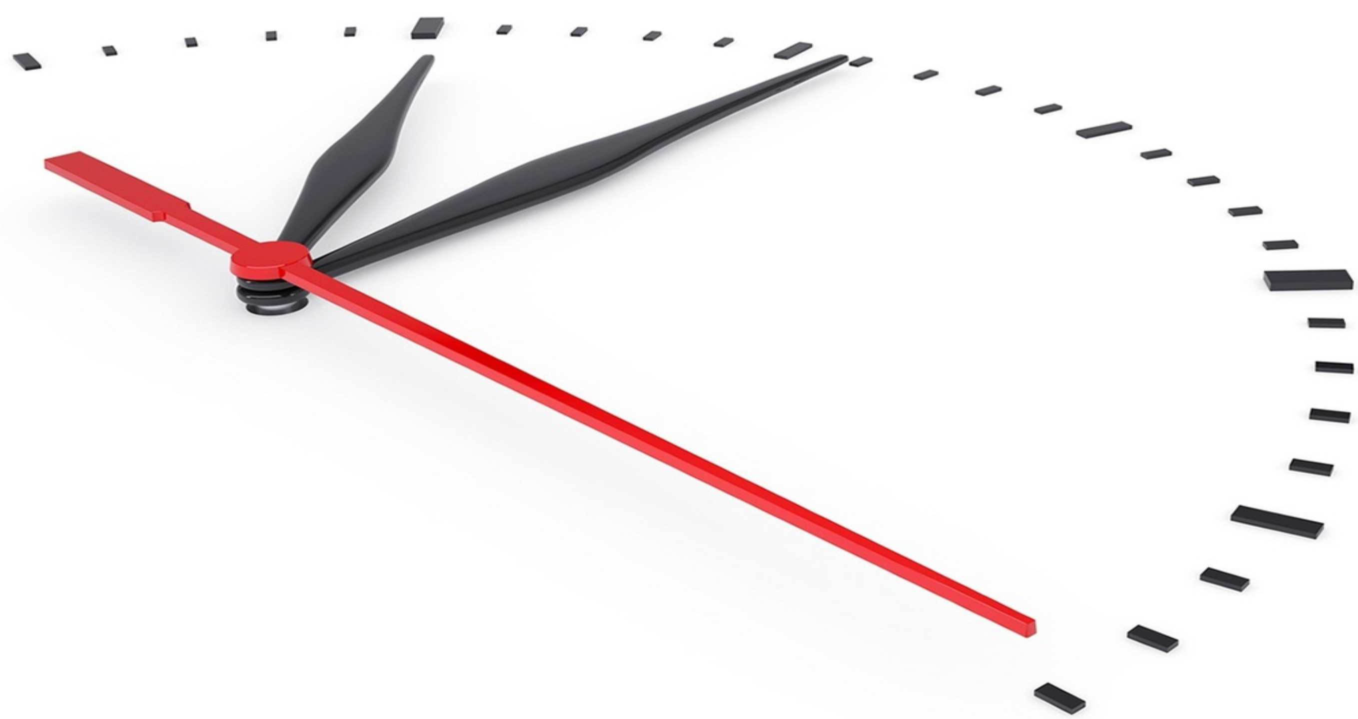
HIDDEN PLAIN SIGHT 3

The secret of time



ANDREW THOMAS

HIDDEN IN PLAIN SIGHT 3

The Secret of Time

Andrew Thomas studied physics in the James Clerk Maxwell Building in Edinburgh University, and received his doctorate from Swansea University in 1992.

His *Hidden In Plain Sight* series of books are science bestsellers.

ALSO BY ANDREW THOMAS

Hidden In Plain Sight
The simple link between relativity and quantum mechanics

Hidden In Plain Sight 2
The equation of the universe

Hidden In Plain Sight 4
The uncertain universe

Hidden In Plain Sight 5
Atom

Hidden In Plain Sight 6
Why three dimensions?

Hidden In Plain Sight 7
The fine-tuned universe

Hidden In Plain Sight 8
How to make an atomic bomb

HIDDEN PLAIN SIGHT 3

The Secret Of Time

ANDREW THOMAS

AGGRIEVED
CHIPMUNK
PUBLICATIONS

AGGRIEVEDCHIPMUNK.WORDPRESS.COM

HIDDEN IN PLAIN SIGHT 3

The Secret of Time

CONTENTS

PREFACE

In its 25th anniversary edition in 2013, *Physics World* magazine published its choice of the five biggest unanswered questions in physics. Its choices were:

- Can we unify quantum mechanics and gravity?
- What is the nature of the dark universe?
- What is time?
- Is life on Earth unique?
- Can we exploit the weirdness of quantum mechanics?

In my first book, *Hidden In Plain Sight*, the first question was considered: how might we find a link between quantum mechanics and relativity? In my second book, imaginatively titled *Hidden In Plain Sight 2*, the second question was considered: what is the nature of dark energy? In this latest book, we will consider the third of these mysterious questions: what is time?

Many books about the nature of time tend to be rather philosophical, and tend to dissolve into endless wordy pages which get us no further along the path of understanding. The problem seems to be that we are too close to the subject: whenever we try to analyse time we inevitably fall back on our own human intuition and feeling about the passage of time. It is very hard to be objective. This book takes a different approach. Physics will be the only guide. We will gain understanding only by analysing the data in an objective manner. The close connection between time and space will prove to be especially useful in our analysis.

In many ways, the book reads like a murder mystery whodunnit. There are plenty of clues as to the secret of time sprinkled throughout the book. But don't be fooled — there's a twist in the final chapter!

Andrew Thomas
(hiddeninplainsightbook@gmail.com)
Swansea, UK
2014

I now have a Twitter account on which I will post updates: twitter.com/andrewthomas101

1

INTRODUCTION

The date: October 1971.

Pan Am flight 106 from Washington's Dulles International Airport is taking off on its scheduled flight to London. To all outward appearances, this was just like any other flight. Passengers were fastening their seatbelts and preparing for the eight-hour trip as the flight stewardesses gave their safety drill. However, we can only imagine the puzzled looks on the faces of some of the travellers as they observed the extraordinary hand luggage of two of their fellow passengers.

For assistant professor of physics Joe Hafele and his colleague, the astronomer Richard Keating, this flight was to form a unique experiment. On the two seats in the middle aisle next to Joe Hafele, no passengers were seated. Instead, the seats were occupied by large cases of electronic equipment, approximately one metre high. These cases contained four highly-accurate caesium atomic clocks.

As part of an experiment, these clocks were being flown around the world. This was to be the first time that the effect of Einstein's theory of special relativity was to be measured using actual clocks. This experiment was going to reveal the true nature of time.

The following photograph shows Hafele and Keating with their clocks onboard the Boeing 747:

Hafele and Keating were attempting to test one of the great insights into the nature of time. For many centuries, philosophers had wondered about the nature of time, without making much progress. The main problem is that the nature of time is tied so closely to our own feelings of the passage of time that it is hard to move away from an intuitive, subjective notion of time to a more objective analysis. We all have our own internal notions of what constitutes time, but it is almost impossible to explain it. Saint Augustine described this dilemma when he said: "What then is time? If no one asks me, I know. If I wish to explain it to one that asketh, I know not."

In his *Critique of Pure Reason*, published in 1781, Immanuel Kant continued Augustine's theme by suggesting that space and time did not exist independently but were constructed by the human mind in order to make sense of the world around us. After all, if time is just a feeling — as Augustine suggested — then maybe it was all in the mind? The French philosopher, Henri Bergson, even believed that this model of time in the human mind would not be present at birth and would have to be constructed via experience. So Bergson believed a newborn baby would not experience time at all! The baby would have to learn to create its own model of time in its head as it grew up.

These philosophical arguments — which seemed to suggest that time purely existed as a subjective notion in our heads — were swept away by Albert Einstein at the start of the 20th century. Einstein was greatly influenced by the philosopher-physicist Ernst Mach who was an advocate of *logical positivism*. According to logical positivism, physics should only make statements about phenomena which could be directly observed and measured. Using logical positivism as his guide, Einstein simply stated: "Time is what we measure with a clock". According to Einstein, there was no place in physics for philosophical musings about the nature of time — all that was important was what could be measured.

This statement of Einstein is particularly important because the theory of special relativity states that a clock which is moving will appear to run at a slower rate than a clock which is stationary. And, as Einstein stated that "Time is what we measure with a clock", this would appear to indicate that time itself runs slower for an observer who is moving relative to a secondary observer.

So this effect of *time dilation* is what Joe Hafele and Richard Keating were trying to measure on their round-the-world aeroplane journey. As Richard Keating said in an earlier interview: "I don't trust these professors who get up and scribble in front of blackboards claiming they understand it all because I've made too many measurements where they don't come up with the numbers they say. It always seemed to me that the best proof is to measure it."

Joe Hafele had been preparing notes for a physics lecture when he performed a brief calculation which showed that an atomic clock on board a commercial airliner should have sufficient accuracy to reveal the effect of time dilation. In order to perform this experiment, Hafele and Keating flew their four clocks around the world: once in the eastward direction, and then in the westward direction. The values on the clocks at the end of the journey were compared with the reference atomic time scale at the U.S. Naval Observatory.

It was found that the flying clocks lost time (aged slower) during the eastward trip, and gained time (aged faster) during the westward trip. The difference between east and west was due to the rotation of the Earth underneath the aeroplane. The variation in time was exactly as predicted by special relativity.

There was also an effect due to the altitude of the aeroplane. As we shall see later in this book, Einstein's theory of general relativity predicts that time passes faster in a weaker gravitational field — such as in an aeroplane cruising at altitude.

As Joe Hafele explained: "Suppose you lived for 100 years, and you spent your entire life on one of these aircraft flying around the world. You could expect to be younger than a person who did not do that by about one ten-thousandth of a second."

We will be considering special relativity and general relativity and the extraordinary effect of time dilation in detail in Chapter Six of this book.

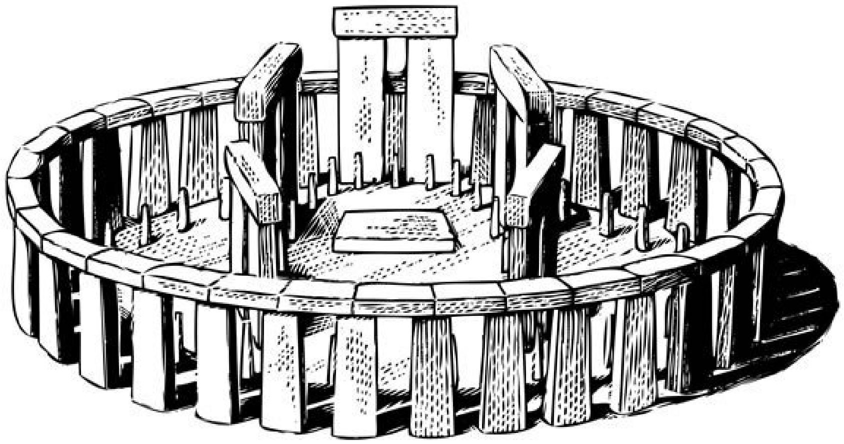
A brief history of time

The very first clocks which were used by early humans were provided by Nature herself. It was clear that the rising and setting of the Sun occurred at regular intervals of one day (we now know this is due to the Earth's rotation on its axis). This regular marking of time provided an easy way to measure periods of time of significant length. The early Egyptians, for example, used the shadows of obelisks as clocks. The passage of the Sun across the sky during the day can also be subdivided into smaller time periods using sundials.

Another celestial clock was provided by the phases of the Moon. The proportion of the Moon illuminated by the Sun varies as the Moon orbits the Earth. The proportion of the Moon which appears to be illuminated when viewed from the Earth varies from 0% (new moon) to 100% (full moon). In between these two extremes we observe the characteristic crescent shapes of the Moon. As the Moon orbits the Earth once every 29.5 days, this regular cycle of lunar phases gave birth to another form of measurement of time: the month.

Finally, the orbiting of the Earth around the Sun — which provides us with the seasons — gave early humans the largest period by which time could be measured: the year. Most famously, the prehistoric standing stones at Stonehenge identify the exact time of the summer and winter solstices. It is easy to forget that one of the functions of Stonehenge was to be one of the earliest clocks.

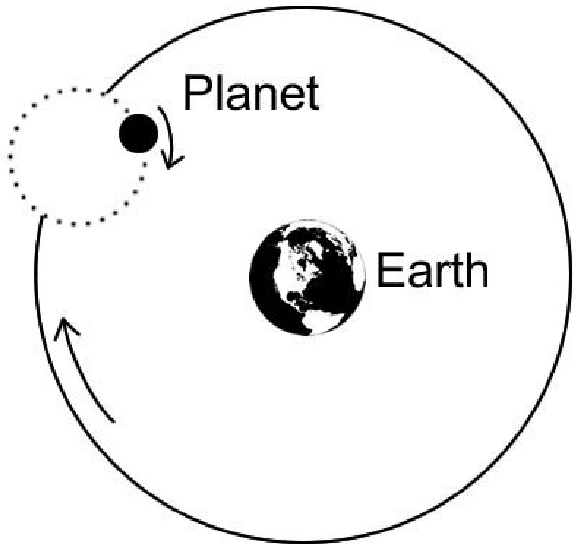
Observe the similarity of the structure of Stonehenge to a clock face:



Ancient Greek astronomers observed the movement of the celestial bodies and tried to model the orbits of the planets using geometry. The Greek image of geometric perfection was the circle, so it was believed the planets should orbit in perfect circles. The Greeks, therefore, modelled the orbits of the planets by a series of concentric rotating transparent spheres, each planet being attached to one transparent sphere. As the spheres rotated, so the planets were observed to move. At the centre of the series of spheres lay the Earth, which was considered to be the centre of the universe. This was, therefore, a *geocentric* model.

However, detailed astronomical observations, such as those obtained by Hipparchus, showed that this geocentric model could not be accurate. The orbits of the planets were more complicated than the motions of concentric spheres. While the Moon, Sun, and stars moved in predictable trajectories, sometimes the planets would appear to reverse their direction in the night sky (so-called *retrograde* motion). In fact, the word "planet" is derived from the Greek word for "wandering star". In order to account for retrograde motion, the astronomer Ptolemy introduced the idea of *epicycles*.

For a planet, an epicycle was a smaller orbit contained within its usual orbit around the Earth. This allowed the planets to sometimes move backwards when viewed from Earth. An example of a planet moving on a epicycle is shown on the following diagram:



This all might appear unnecessarily complicated to our eyes now, but the geocentric model of the universe was generally accepted until the 16th century when the Polish astronomer Nicolaus Copernicus realised that the model could be greatly simplified by placing the Sun at the centre of the Solar System and having the planets (including the Earth) orbit the Sun. This formed the *heliocentric* model.

Copernicus's model was more elegant than the Ptolemaic model. It not only explained the retrograde motions of the planets but it also explained why the Earth experienced the seasons as it orbited the Sun once a year.

In the more general sense, the more recent development of the so-called *Copernican principle* states that no particular point in the universe (not just the Earth) can hold a privileged position in the universe. This heralds a move away from an absolute system of science and cosmology to a science which realises that no observer holds a privileged position, and that the universe is built on relative measures. The repercussions of the Copernican principle are rippling through science to this day, perhaps having its greatest impact in the theory of relativity: if no observer holds a privileged position, then all motion must be described relatively.

If you have read my previous two books you will know I have a firm conviction that the universe is built on fundamental principles — principles which are "obviously correct" and would have to be true in any conceivable universe. The Copernican principle is surely another of these fundamental principles: surely no point — and therefore no observer — holds a privileged position in any conceivable universe. In many ways, as you will see, this is a book which is based on the importance of the Copernican principle and all that it entails. We will encounter the Copernican principle again several times in the later chapters of this book, when it will be shown that it might possibly hold an important key to explaining the nature of time.

These astronomical measurements of time, such as the period of rotation of the Earth, remain extremely accurate measurements of time to this day. In fact, until 1967 the length of the second was defined in terms of the orbital period of the Earth (to be precise, it was defined as a fraction of the time taken for the Earth to orbit the Sun in 1900). However, as society became more complex, more accurate subdivisions of time were required.

In the Middle Ages, slow-burning candles with colour-coded wicks were used to mark time in monasteries. Elaborate water clocks (dripping taps, basically) could be found in wealthy households. However, it was the invention of the mechanical clock which really brought time to the masses. As great cities emerged throughout Europe, together with the rise of commerce and trading, there was a need to find some way to synchronise commercial activities throughout the city. Mechanical clocks were installed in the bell towers at the heart of cities, with the daily activity of the city being based around the various chimes. The first public clock was installed in Orvieto in Italy in 1307, and this innovation spread rapidly throughout the rest of Europe.

But while it was possible to synchronise the activities of a single city by using a bell tower, this was no way to synchronise the activities of multiple cities. Each city could operate to its own local time standard, which might be a completely different standard from that of another city (generally, noon in each city was set to the time when the Sun was at its highest point in that city). This meant even simple activities — such as organising a meeting between citizens of different cities — were fraught with difficulties. This failing became more pronounced with the coming of the industrial revolution.

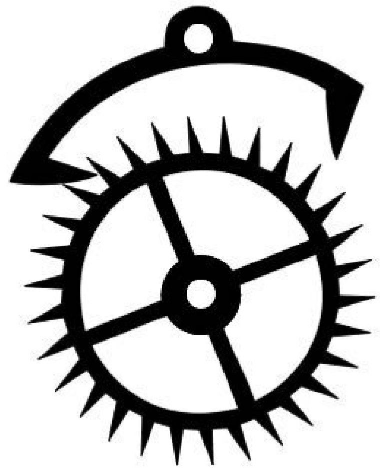
The industrial revolution brought steam-powered railways, which connected many cities in a single network. The emergence of the railways made the introduction of a network-wide time standard essential in order for services to run efficiently (we will see in Chapter Three how this requirement for train synchronisation became a particular interest for a patent clerk working in Zurich).

As an example of the problems which were occurring, in 1841 in England, the Great Western Railway was extended to Bristol. However, passengers arriving at Bristol Temple Meads station were annoyed to find the trains were leaving eleven minutes early. The problem was because the trains were coming from London and were using London time. Bristol is 200 miles west of London so the Bristol sunrise is eleven minutes later than the London sunrise. Hence, Bristol time was eleven minutes behind London time. Some form of national time synchronisation was clearly necessary, requiring the most accurate clocks.

The necessary advance in accuracy had been provided many years earlier by Galileo. Supposedly, Galileo was in Pisa Cathedral when he observed the swinging motion of a large bronze lamp. The ever-inquisitive Galileo timed the swings using his pulse and found that the period of the swings was always the same — no matter how big was the length of the swing. So as the lamp lost energy, and the amplitude of the swing decreased, each swing still took the same amount of time. Galileo showed that the time taken for a complete swing of a weight on a pendulum was independent of both the size of the weight and the length of the swing, and was only dependent on the length of the pendulum (a pendulum one metre long swings once every second — the usual length of the pendulum in a grandfather clock).[1] This allowed for great accuracy if the length of the pendulum was measured correctly. Previous mechanical clocks were only accurate to about 15 minutes a day. Pendulum clocks were accurate to about 15 seconds a day.

Pendulum clocks remained the most accurate standard for timekeeping for 270 years until the invention of the quartz clock in 1927. The most accurate pendulum clock, used as the US standard time until 1929, utilised a pendulum in a partial vacuum and had an error of only 10 milliseconds per day.

The great innovation of mechanical clocks was the *escapement*: a device which rocked backwards and forwards over a toothed wheel. The escapement allowed for the controlled release of energy at a regular rate. It consists of a toothed wheel with a rocking lever above it. As the lever rocks backwards and forwards, the toothed wheel advances by just one notch. This is responsible for the characteristic clockwork "ticking" sound:



This principle of the "tick" of a clock — the regular marking of small periods of time — is worth examining in detail. In this chapter we have considered a wide range of different types of clocks, from the Sun passing across the sky, to a swinging pendulum, and in the next section we will be considering the world's most accurate clock: the atomic clock. But there is one thing which all these different types of clock have in common, and that is that they all have a "tick". A tick is an oscillation — a recurring event — which must happen at a regular time interval. It is easy to see that a pendulum swings at a regular time interval, but even the Sun passing across the sky represents a "tick" which occurs at a regular time interval: once a day.

Ticks can be counted, and the result displayed on the output of the clock in order to produce a measurement of time.

But why is the regularity of the tick so important? You might feel this is a trivial and obvious question, but it is actually an important question whose answer leads us to an important insight into the nature of space and time.

Firstly, let us consider the implications for the nature of space.

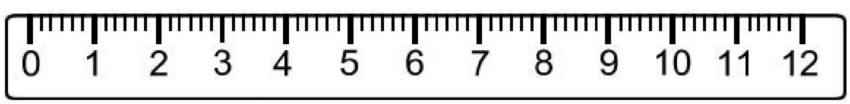
If we perform an experiment, and make a note of the result, and then we move the experimental apparatus six feet to the left and perform the identical experiment again, we find we will get the same result. Likewise, if we measure the width of an object using a ruler, and then move the object twenty feet to the right and measure it again using the same ruler, we find we will get the same measurement for the width of the object. This principle that the laws of Nature work in exactly the same way no matter where the experiment is performed is called *space translation invariance*.

Now consider we perform the same experiment again, but instead of performing the experiment at a different location, we perform the experiment at a different **time**. As long as the experiment is identical, we find we will get the same result. Likewise, if we measure the duration of an event using a clock, and then we perform the identical experiment again at a later date we will find the duration of the event will be exactly the same. This principle that the laws of Nature work exactly the same no matter **when** an experiment is performed is called *time translation invariance*.

Because of space translation invariance, we must ensure that any measuring equipment we use to measure distances (e.g., a ruler) must also possess space translation invariance, i.e., it does not matter if we shift our ruler left or right — it will still give the same measurement. This means that the marks — the "ticks" — on our ruler must all be equally spaced.

And because of time translation invariance, we must ensure that any measuring equipment we use to measure time (e.g., a clock) must also possess time translation invariance, i.e., it does not matter if we perform the experiment sooner or later — the clock will give the same time measurement. So it is vital that the "ticks" of our clock are equally spaced and occur at regular time intervals because of time translation invariance.

If we now consider the regular arrangement of "ticks" on a 12-inch ruler:



and we bend that ruler around in a circle:

It becomes a clock face!

Space translation invariance results in equal ticks along a ruler, and time translation invariance results in equal ticks around the face of a clock.

This is not a trivial result. This similarity gives us the first hint of a deep connection between space and time, a connection which we will be exploring throughout this book.

The world's most accurate clock

Atomic clocks are the most accurate clocks ever built. As we have just discussed, any clock requires some regular oscillation ("ticks") which form its time standard. In an atomic clock, the time standard which is used is the time it takes for an electron to jump between orbits (energy levels) inside an atom. When electrons jump to a lower energy level, they produce electromagnetic radiation. This radiation might be in the microwave, optical, or ultraviolet region. The pattern of frequencies at which energy is emitted is called the *spectrum*. Each chemical element has a characteristic spectrum. For example, we can deduce the component elements of a star merely by looking at its light spectrum.

For a particular element, one of the frequencies in its spectrum can be selected as the "tick" of an atomic clock. As an example, one of the frequencies of the element caesium is frequently used in atomic clocks (remember, the atomic clock used in the Hafele-Keating aeroplane experiment was based on caesium). This "ticking" forms a particularly reliable time standard. Since 1967, the International System of Units (SI) has defined the second to be 9,192,631,770 ticks of a caesium atom (superseding the previous definition based on the orbit of the Earth around the Sun).

The most accurate atomic clocks in existence could have been used to measure the age of the universe to an accuracy of one second.

An interesting thought now might occur to us: we measure the accuracy of a clock by comparing its measurement with a more accurate clock. But atomic clocks are the most accurate clocks in existence. How can we possibly measure the accuracy of the world's most accurate clock? After all, we cannot determine its accuracy by comparing it with any other clock which is more accurate — no such clock exists!

Dr. Stefan Droste of the Max Planck Institute of Quantum Optics is part of a team which is developing a highly-accurate atomic clock in the optical range of frequencies (most atomic clocks use lower microwave frequencies).

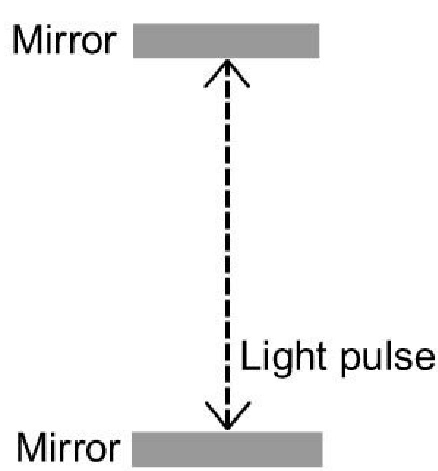
As Dr. Droste explains: "Time, as we know it, is based on caesium atomic clocks. These kinds of clocks have accuracies to the order of 1×10^{15}. The confidence in the time that clocks show is increased by comparing hundreds of clocks around the world with one another, and this is important since with only one single clock, you cannot know whether your clock is showing the correct time or not." [2]

So the answer as to how you measure the accuracy of the world's most accurate clock is that you build another clock — or a series of clocks — exactly like the first clock. You then run all of these clocks for a period of time and, at the end of that period, you see if the time measured by the clocks has varied from clock to clock. Though the variation will be small, it will not be zero — it can be measured. And that variation gives you a value for the accuracy of that particular type of clock over time.

This is essentially what Hafele and Keating did when they flew four atomic clocks around the world rather than just a single clock. By using multiple clocks it is possible to obtain a value for variation of accuracy in the experiment.

No atomic clock will ever be 100% accurate: there will always be some drift. The only truly accurate measure of time can be produced by a clock whose "tick" is based on the speed of light. Such a clock is called a *light clock*, and Einstein considered such a clock as the basis for his groundbreaking 1905 paper on special relativity.

The light clock is simply composed of two perfect mirrors which face each other, with a pulse of light bouncing between the two mirrors:



Each time the pulse bounces between the two mirrors represents one "tick" of the light clock. Unfortunately, no practical light clock has ever been built. If such a clock could ever be perfectly manufactured, it would instantly be recognised as the most accurate clock in existence. This is because the accuracy of that clock would be entirely based on the speed of light, and the speed of light is a fundamental physical constant the value of which is known not to vary with time.

However, the question then arises: how do we measure the accuracy of the light clock? How can we be absolutely certain that its accuracy does not drift over time? There would be no point using any other type of clock — a caesium atomic clock, for example — to measure its accuracy as no other clock could be as accurate as the light clock. The light clock, therefore, represents a time standard which can be unmatched by any other type of clock. The light clock is the clock by which all other clocks are judged.

Effectively, this means that all other clocks — every atomic clock, the clock by the side of your bed, your wristwatch — are subservient to the light clock. It is the light clock which decides if your wristwatch is running slow — there is no better standard than the light clock.

So instead of time defining the behaviour of a clock, the light clock essentially defines time!

This raises an interesting point because, as we shall see later in this book, the behaviour of a light clock is subtly different for a moving observer. So if a light clock effectively defines time, then **time itself must be modified for a moving observer!** This is precisely what Hafele and Keating confirmed in their round-the-world experiment. We will be returning to consider this point in detail later in the book.

But what is so special about the speed of light?

2

THE SPEED OF LIGHT

The next two chapters are going to describe a single, long journey. It is a journey which will commence at the very start of science. It is a journey which will take four hundred years and involve several of the greatest physicists in history. It is a journey which will end with an experiment which, if correctly analysed, will change everything you ever believed about reality.

However, every journey begins with a first step …

The Galilean transformation

Galileo Galilei was born in Pisa in 1564. His father was a famous musician, and this highly-cultured family environment provided Galileo with the perfect opportunity to explore his talents. He was taught at the University of Pisa where he became fascinated by physics and mathematics.

In 1609, Galileo was in Venice when he heard of a new invention which allowed distant objects to appear as though they were close. Galileo managed to improve on the design of the invention and invited a group of merchants to climb St. Mark's bell tower for a demonstration. When Galileo unveiled his telescope, the merchants were delighted as it allowed them to spot trading ships arriving over the horizon forty miles away. As Venice was such a commercial hub at the time, this gave them a considerable advantage over their rivals.

The scene is shown below in a fresco by Bertini:

Galileo made a significant amount of money from the telescope, and continued to invent throughout his life. But Galileo's greatest contribution to the world of science was his *scientific method*. Before the scientific method was introduced, the proclamations of ancient philosophers such as Aristotle were accepted without question as representing the absolute truth about Nature. Galileo was one of the first scientists to challenge the wisdom of the ancients. Galileo's introduction of the scientific method allowed Nature to speak for itself.

Aristotle was no experimenter, and he relied too much on his preconceptions. Famously, Aristotle once proclaimed that women have fewer teeth than men. Because no one thought to check this proclamation of Aristotle, for a thousand years everyone believed that women have fewer teeth (women and men, of course, actually have the same number of teeth).

Aristotle's poor method of investigation was described by Bertrand

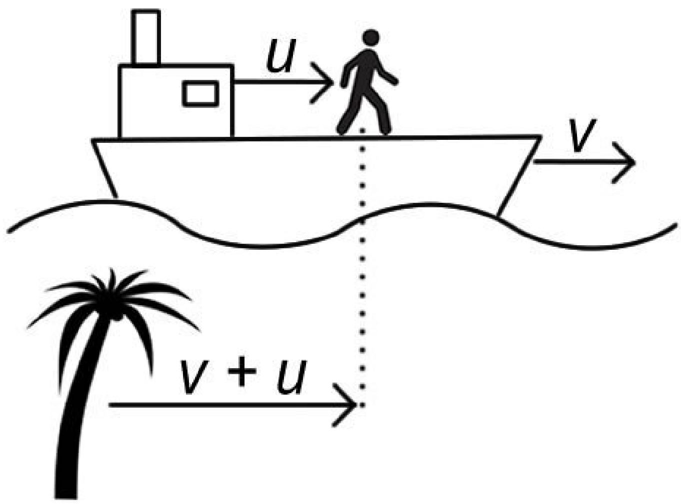
Russell: "Aristotle maintained that women have fewer teeth than men. Although he was twice married, it never occurred to him to verify this statement by examining his wives' mouths."

Galileo's scientific method replaced these dubious proclamations. Firstly, according to the scientific method, a hypothesis is proposed about some testable aspect of the natural world. Secondly, that hypothesis is checked through observation and experiment. Based on the result of the experiment, the hypothesis is either supported or rejected. Galileo performed many such experiments using apparatus such as wooden wedges and balls for testing his theories of motion. For introducing the world to the scientific method, Galileo is often referred to as the father of modern science.

However, Galileo got into trouble when he turned his telescope toward a wider horizon. The discovery of the four moons orbiting Jupiter — Io, Europa, Ganymede, and Callisto — suggested that the Earth was not the centre of the universe about which all celestial bodies orbited. By challenging the geocentric model of the Solar System, Galileo found himself accused of heresy and was placed under house arrest for the rest of his life.

It was while under house arrest that Galileo published his final book entitled *Discourse on Two New Sciences* (the two sciences being mechanics and the strength of materials). The book was the culmination of Galileo's work over the preceding forty years. We are going to be considering two of the principles of motion described in the book. The first principle we will consider deals with the addition of velocities.

As Galileo was living in a world in which the ship was the dominant form of transport, he devised a maritime-themed explanation of his principle. Consider the following diagram. Imagine a ship is moving at a speed of v, and a man is walking on the ship at a speed of u (relative to the ship). Then from the shore (denoted by the palm tree) the speed of the walking man will be measured to be the sum of v and u:



This addition of velocities is called the *Galilean transformation*.

This principle is every bit as simple and intuitive as it might appear at first glance, but it took the innovative mind of Galileo to publish the principle for the first time. What is important to note is that when an object is emitted by a source — for example, a man firing an arrow from a bow — you have to add the velocity of the source to find the final velocity of the object. This might seem trivial to our eyes, but it represented a great insight in the 17th century.

The second principle of Galileo we will consider is the principle of *Galilean relativity*. Put simply, Galileo stated that if you were travelling in a straight line at a constant speed you would feel as if you were stationary. To be more precise, there would be no experiment you could perform which could reveal if you were either stationary or moving at a constant velocity. For example, if you are sitting in an aeroplane and you throw a ball straight up in the air, the ball will come straight back down into your hand. Even though the aeroplane is moving extremely fast, the ball will not rush to the back of the plane. So to all intents and purposes it feels as if you are stationary. Hence, Galilean relativity states that the laws of motion are the same for any observer in uniform motion.

We will be returning to consider this principle of Galilean relativity in the next chapter when we consider how Einstein imagined running alongside a ray of light.

Measuring the speed of light

Galileo performed an experiment for determining the speed of light by having two observers with lanterns face each other over a great distance. The lanterns had shutters which could be quickly opened and closed. When one of the observers saw a flash of light from the distant lantern, he had to reply by quickly flashing his own lantern. This rate of flashing of lanterns could be timed and, when combined with the distance between the two lanterns, it should have been possible to calculate a value for the speed of light.

However, when Galileo performed the experiment he found the time taken for the light to travel over a mile distance was virtually instantaneous. If the speed of light was not actually instantaneous then it certainly had to be extremely fast.

In 1676 the Danish astronomer Ole Romer showed that light does not travel instantaneously but instead has a finite speed. He obtained his measurement by observing the eclipses of Io — one of the moons of Jupiter. He noted that the time of the eclipse became earlier as Earth approached Jupiter, and became later as Earth moved away. Romer explained the difference in the eclipse timings as being due to the additional time which light took to reach the Earth. Romer presented his result to the French Academy of Sciences, announcing that the speed of light was 220,000 kilometres per second (this is actually 26% less than the true value which we now know is 299,792 kilometres per second).

In 1850, the French physicist Leon Foucault bounced a beam of light off a rotating mirror to a distant fixed mirror. When the beam of light reflected back to the rotating mirror, the mirror had rotated by a certain known amount. Hence, the light came off at a slightly different angle. From knowing the speed of rotation of the mirror and the angle at which the light was reflected, Foucault measured the speed of light to be 299,796 kilometres per second, which is just 0.001% greater than the actual value.

Maxwell's equations

Our journey now moves in a surprising direction, to consider the science of electricity and magnetism: *electromagnetism*.

Michael Faraday was the self-taught son of a blacksmith, but became an experimental physicist of the highest order. In 1831 in London, Faraday demonstrated to the Royal Society that a magnet moving through a coil of wire produces an electric current in that wire. This was the discovery of *electromagnetic induction*, and it is the same principle which allows modern power stations to produce electricity. The principle that a changing magnetic field produces a changing electric field is called *Faraday's law*.

Faraday's result seemed to be the inverse of a earlier result which was obtained by the French physicist Andre-Marie Ampere. Ampere showed that an electric current through a wire created a magnetic field. This became know as *Ampere's law*.

But it took a great mathematical physicist to combine these two results of Faraday and Ampere, thus revealing the remarkable underlying significance.

At the end of the 20th century, *Physics World* magazine conducted a poll of 100 leading physicists to discover who was considered to be the greatest physicist of all time. There was no surprise when Albert Einstein and Isaac Newton topped the poll, but in third place came a physicist who is largely unknown to the general public. His name was James Clerk Maxwell, and there is no doubt as to the esteem he is held in by physicists. According to Max Planck: "He achieved greatness unequalled." According to Einstein: "The work of James Clerk Maxwell changed the world forever." And according to Richard Feynman: "From a long view of the history of mankind — seen from, say, ten thousand years from now — there can be little doubt that the most significant event of the 19th century will be judged as Maxwell's discovery of the laws of electrodynamics."

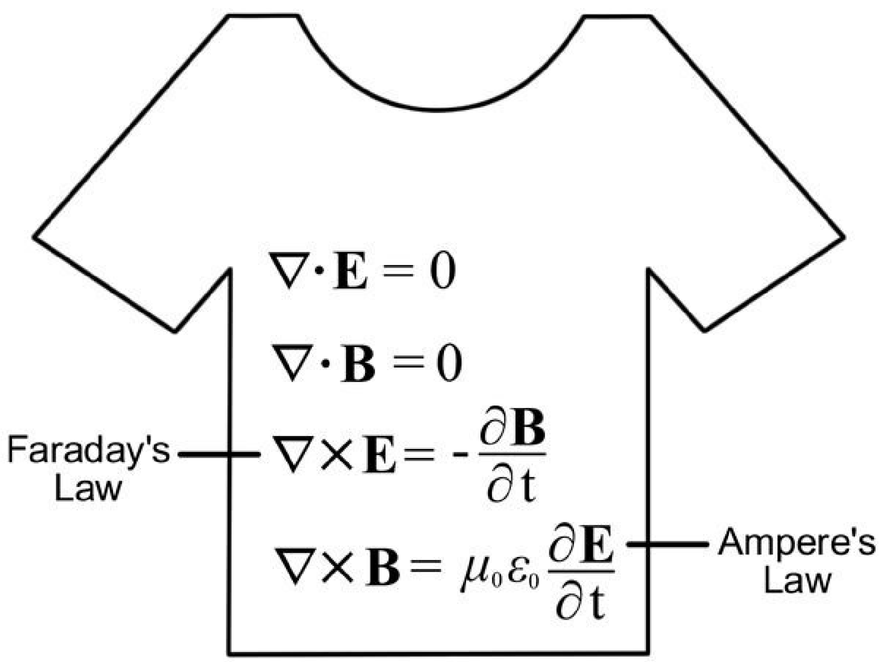
James Clerk Maxwell was very much the archetypal Victorian gentleman.

He was born in 1831 into a God-fearing, hard-working Edinburgh family. Maxwell displayed an early talent for mathematics and geometry, and earned a place as an undergraduate at Edinburgh University.

Maxwell's stature as a mathematician grew rapidly in the academic environment. He also possessed an inquiring mind for many aspects of general science (what was then called "natural philosophy"), and in his spare time performed many experiments which intrigued him. He was particularly interested in the properties of light. By using a spinning colour wheel he was able to show that white light could be formed by a combination of red, green, and blue light. Maxwell's principle was used to create the world's first colour photograph in 1861, produced from a set of three separate monochrome images, and this was used to illustrate a lecture on colour by Maxwell.

Maxwell would regularly attend lectures at London's Royal Institution where he came into contact with Michael Faraday. Faraday had great experimental technique, but weak mathematical skills. Maxwell was able to use his own considerable mathematical skills to cast Faraday's Law into mathematical form, and to combine it with the result of Ampere.

As a result, Maxwell created four equations which, rather ironically, are now probably best known for being a very popular T-shirt design:

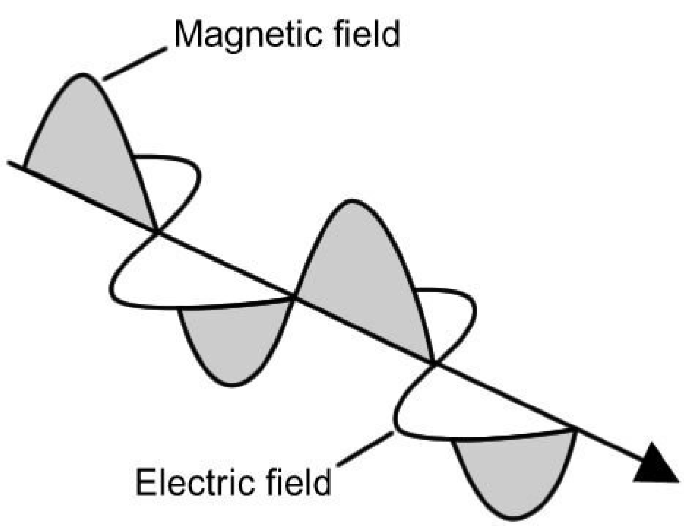


We are only interested in the last two equations which represent Faraday's law and Maxwell's form of Ampere's law. The equation for Faraday's law states how the electric field, *E*, is generated by a changing magnetic field, *B*. And the equation for Ampere's law states the inverse: how the magnetic field, *B,* is generated by a changing electric field, *E*.

Maxwell's great stroke of genius was to realise that these two equations had the same form as the equation for a wave travelling through space. To see this, remember once again that the equations give us the following symmetrical results:

1. A changing magnetic field creates an electric field.
2. A changing electric field creates a magnetic field.

Hence, the magnetic field can generate the electric field, and the electric field can generate the magnetic field, and so on forever. The result is a self-sustaining wave of alternating electric and magnetic fields called an *electromagnetic wave*:



Maxwell realised that, according to his equations, the speed of the resultant wave could be determined by two electric and magnetic constants which had been determined by experiment. Hence, the speed of the wave, *c*, is given by:

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

where the magnetic constant, μ_0, has a numeric value of $1.25663706 \times 10^{-6}$, and the electric constant, ε_0, has a numeric value of $8.85418782 \times 10^{-12}$. If you substitute these numeric values into the above formula (try this at home with your calculator), you will find you get a value for *c* of approximately 299,796,000 m/s. Incredibly, Maxwell realised that this was exactly the same speed as Leon Foucault had measured for the speed of light! So Maxwell made the inspired leap to deduce from this that light was a form of electromagnetic wave.

Visible light is not the only form of electromagnetic radiation. Depending on the wavelength, other types of electromagnetic radiation include radio waves, microwaves, infrared, ultraviolet, X-rays, and gamma rays. By discovering the fundamental form of all these different types of electromagnetic radiation, Maxwell brought about the development of radio and television, as well as many other modern technologies. According to Carl Sagan: "Maxwell's equations have had a greater impact on human history than any ten presidents."

But Maxwell's equations predicted some surprising properties of electromagnetic waves. Firstly, the equations as printed on the T-shirt in the earlier diagram are the equations for radiation through a vacuum containing no electric charges or electric currents (if electric charges or currents are present then the equations become more complicated). This is fine for our purposes as we are interested in the speed of light in a vacuum. However, this confused physicists at the time. All known waves up to that time were a form of disturbance in some underlying medium. For example, sound waves were a disturbance in air, and sea waves were a disturbance in water. But Maxwell's equations seemed to indicate that electromagnetic waves could travel through a vacuum containing no medium which could carry the waves. All experiments to detect an underlying substance — called the *ether* — drew a blank.

We now know that the ether does not exist, and electromagnetic waves can, indeed, travel through a total vacuum. After all, how else could light reach us from the stars across the intergalactic void?

It is interesting, at this point, to digress slightly and tell the story of what happened when I gave this book to a friend of mine for proof-reading. He took great exception to my description of the self-sustaining wave travelling through space, with the electric and magnetic fields oscillating off each other. His objection was based on the fact that it sounded like a form of perpetual motion and, as he stated with great certainty, there was no such thing as perpetual motion. Well, I had to break the bad news to my friend that this was, indeed, a form of perpetual motion and it is the reason why light from the stars can reach us across billions of light years. Indeed, if the light was not intercepted by us it would continue on its merry way across the universe. So this is most certainly a form of perpetual motion — we do not see light slowing down! We do not see light running out of energy and coming to a stop. So the next time a man down the pub tells you that perpetual motion is impossible, just ask him when was the last time he saw a ray of light coming to a halt?

In fact, one of the themes of this book, which we examine in detail in the later chapters, is that if we are to achieve a fuller understanding of time then we will have to overcome our in-built bias against the principle of perpetual motion. As we will find, perpetual motion is all around us.

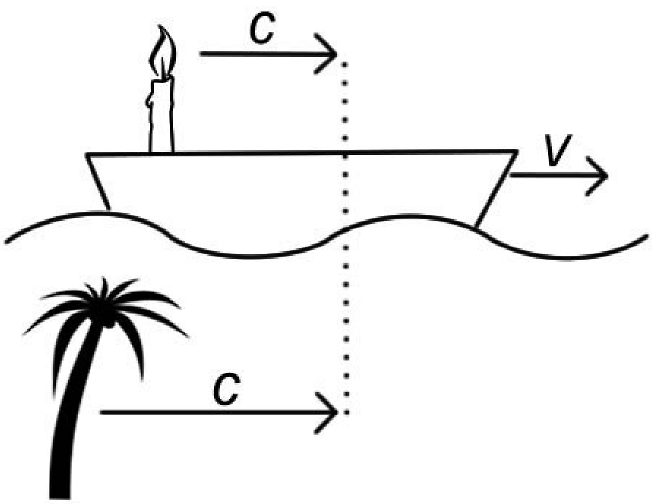
But, to return to the surprising properties of electromagnetic waves, let us consider again the formula for the speed of the wave:

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

What is so truly astonishing about this result is that **it contains no term representing the speed of the source!** Remember back to our discussion of Galileo's principle of addition of velocities earlier in this chapter. It was explained how the velocity of the source would always have to be added to the velocity of an object to calculate the final velocity. However, Maxwell's formula for the speed of light does not say that the speed of the source has to be added. It is as if every observer would measure the same speed of light —

regardless of the motion of that observer.

To show how bizarre this result really is, let us replace the man walking on the ship with a ray of light being emitted from a candle on the ship:



On the basis of Galileo's principle for the addition of velocities, we might expect the observer on the shore to measure a greater value for the speed of the ray of light than the ship-borne observer (the speed of light measured by the observer on the shore would have the speed of the ship added to it). However, as shown on the diagram, the result of Maxwell would appear to indicate that the observer on the ship and the observer on the shore would both measure the same value for the speed of the light emitted by the candle (i.e., the observer on the shore does not have to add the speed of the ship).

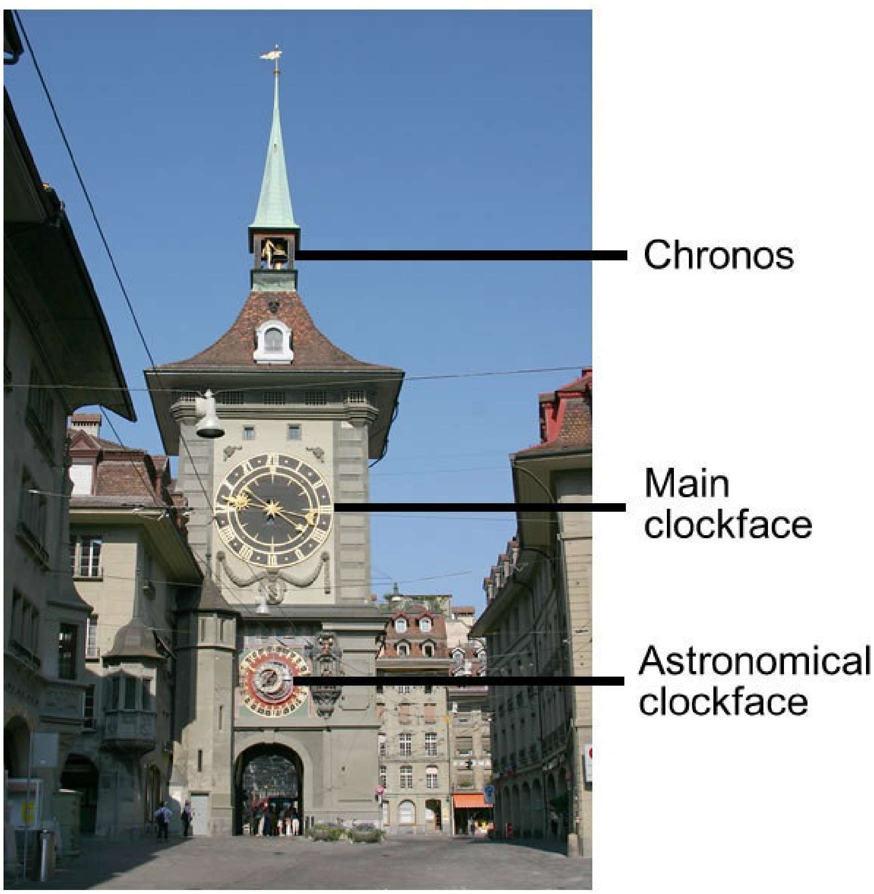
This result seems to contradict everything we intuitively understand about motion. But what are its full implications for reality?

3

SYNCHRONICITY

The date: June 1902.

Albert Einstein has moved to Bern having just accepted a job at the patent office. Together with his wife and young son he is living in an apartment at No. 49 Kramgasse (see http://tinyurl.com/einsteinhouse — the house is now preserved as a museum). Every morning when he left his apartment to go to work he would have turned left and walked past the world-famous clock tower at the end of his street.



As shown on the previous image, the clock tower displays a welter of temporal information. Underneath the main clock is an astronomical clock which displays the phases of the Moon. It also has a model of the Sun which moves around the dial to indicate the current time of sunrise and sunset. Every hour, a clockwork-powered automaton emerges at the top of the tower to strike the bell. This is the gilded figure of Chronos, the Greek god of time.

In 1902, the clock tower played a particularly important role as it was used as the official time keeper for the new railway station. Effectively, it supplied the official time standard for the city of Bern. Other cities were not necessarily synchronised with the time of Bern, so as trains passed the station at speed they would synchronise their clocks with the time seen on the clock tower.

As Einstein took his desk as a Technical Expert Class 3 at the patent office, he would use his considerable technical ability to consider the virtues and originality of the day's various patent applications. According to Walter Isaacson in his biography of Einstein, these patent applications included "dozens of new methods for synchronising clocks and coordinating time through signals sent at the speed of light."

From the complex and fascinating clock tower he passed each morning, to the various ingenious patent applications he dealt with in his job, Einstein was surrounded by stimulating aspects of time and synchronisation.

Einstein was also well-informed about the latest developments in physics. When he was an undergraduate at the Zurich Polytechnic he became increasingly frustrated that the latest results about electromagnetism proposed by James Clerk Maxwell were not studied as part of the curriculum. This was because, since the age of 16, Einstein had imagined riding alongside a beam of light. What would you see? Would the light appear to be stationary? As we have seen, the latest result of Maxwell seemed to indicate that, no, the light would not appear to be stationary — you would still measure the same speed of light.

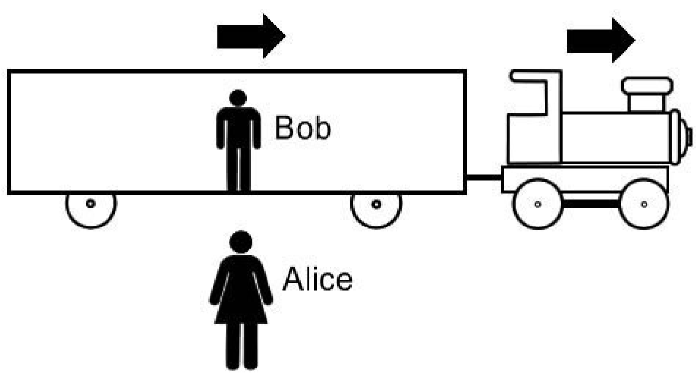
Einstein accepted the result of Maxwell — no matter how bizarre it seemed. If you remember back to the discussion of Galilean relativity in the previous chapter, you will remember how Galileo stated that the laws of motion were the same for all observers in uniform motion. By accepting that the speed of light was the same for all observers in uniform motion, Einstein extended Galilean relativity to cover all of the laws of physics — including the laws of electromagnetism. This became the *principle of relativity*: **the laws of physics are the same for all observers in uniform motion.**

I think the reason this principle might seem fairly reasonable to our eyes is because it seems a logical extension of the Copernican principle which was considered in Chapter One. The Copernican principle states that no point — and no observer — in the universe holds a privileged position. It certainly seems reasonable to assume that the laws of physics apply equally to all observers — the universe is a very equitable place.

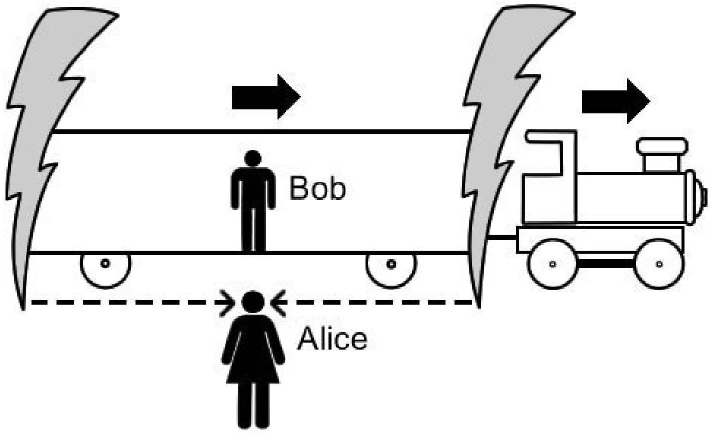
As part of Einstein's job at the patent office was to analyse patent applications to determine if they would be successful in the real world, a desk-bound Einstein would have had to construct imaginary "thought experiments" in his head to examine their potential. Einstein now applied this technique to analysing the implications of this observer-independent speed of light.

The thought experiment which will now be described is a mainstay of popular science books. We will be considering the experiment in more detail than usual as I believe most descriptions of the experiment do not describe it in sufficient detail. As a result, the essential extraordinary conclusion of the experiment is often missed.

We will consider a train travelling along a track at a constant speed. The train has a single carriage. Inside the carriage is Bob, who is standing in the middle of the carriage. Outside the carriage is Alice, who is standing on the station platform watching the train as it rushes by:



Next, something remarkable happens. At the brief moment when Bob is directly in front of Alice, Alice sees two bolts of lightning hit the track at the same time at precisely the position of the front and back of the carriage. The bolts conveniently leave marks on the track (so the positions can be found later), and they also leave marks on the front and back of the carriage:



Alice assumes the lightning bolts have hit the track at exactly the same time because she sees the light from the bolts reaching her at the same time (see the dashed lines on the previous diagram). But, from this observation, can Alice necessarily be certain that the lightning bolts actually hit the track at the same time? Alice realises that there are two factors which could have influenced this result and caused the light from the two bolts to reach her at the same time — even if the bolts did not hit the track simultaneously. Firstly, the distance from her position to each of the two lightning bolts was maybe not equal. So, to check this, Alice measures the distance between her position and the two marks on the railway tracks which were left by the two bolts. Alice finds the distance to each of the two marks is exactly the same. So that part of the experiment is satisfactory.

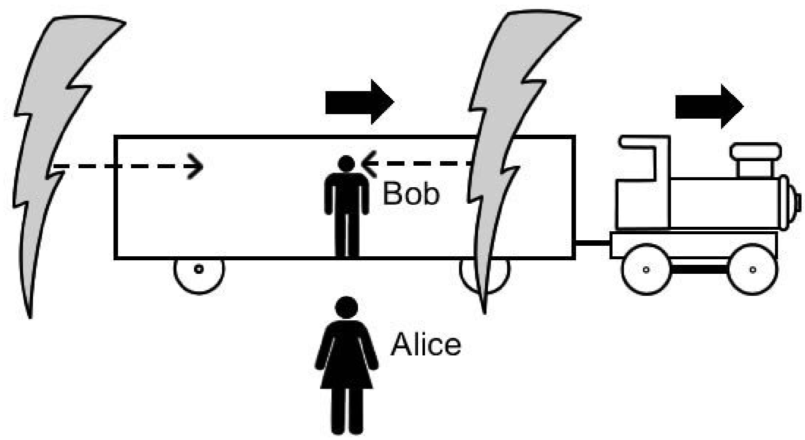
Secondly, Alice realises that maybe the speed of the light was not the same from each of the two bolts. Any inequality in the speed could have resulted in the light from the two lightning bolts reaching her at the same time — even if the bolts did not hit the ground simultaneously. However, Alice had installed equipment to measure the speed of light from each of the two bolts. Alice finds that, just as James Clerk Maxwell predicted, the speed of light was exactly the same from each of the two bolts.

So Alice is left with a surefire conclusion, a logical certainty: the two bolts really hit the ground simultaneously. The distance was the same, the speed of light was the same, and the light from both bolts reached her at the same time. It is therefore an absolute fact: both lightning bolts hit the ground at the same time.

Now let us turn our attention to Bob who is riding on the train. As the train continues its progress down the track at a constant speed, Alice observes

Bob's situation. Bob continues to get closer to the point where one of the lightning bolts hit the track by the front of the carriage. Hence, he moves closer to the light coming from that point in front of him. In the process, he moves further away from the point on the track behind him where the other lightning bolt hit the track.

So in the next diagram we see the train which has moved to the right, and so Bob has moved closer to the lightning bolt near the front of the train:



Hence, the light from the lightning bolt in front of Bob reaches Bob first. Alice and Bob both agree on this fact.

So the light rays from the two bolts do not reach Bob simultaneously. But, from this, should Bob conclude that the lightning bolts did not strike the ground simultaneously? Of course not. Bob is no fool. Bob suspects this is merely an observational anomaly. He realises the train is moving. He realises the movement of the train is taking him nearer to the position of the front lightning bolt. Surely that is the only reason that light from the front lightning bolt reached him first. Bob realises that there is only one way to calculate if the lightning bolts really hit the ground simultaneously, and that is to perform a few measurements.

So firstly, just like Alice did, Bob measures the distances. Remember that the two lightning bolts conveniently left marks at the front and the back of the carriage. So Bob measures the distances from his standing position to the front and the back of the carriage and he finds both distances are the same. So that is certainly not the reason why light from the front bolt reached him first.

Bob is not too surprised about that, because Bob really suspects that it is the apparent speed of the light which is the culprit. After all, he knows he is on a moving train. His movement toward the position of the front lightning bolt would have surely resulted in a greater effective speed of light from that bolt. Fortunately, just like Alice, Bob had some accurate equipment installed for measuring the speed of light from both lightning bolts.

However, when Bob measures the speed of light from both bolts **he finds the speed is the exactly same!** This seems bizarre, but it is precisely in accordance with the prediction of James Clerk Maxwell who stated that all observers will measure the same speed of light. So Bob is left with something of a mystery, and he examines the evidence:

1. The distance to each source of light at the front and the back of the carriage is exactly the same.

2. The speed of light from each light source to Bob is exactly the same.

3. The light rays did not reach Bob simultaneously.

Bob is clearly left with only one logical conclusion: **the lightning bolts did NOT hit the ground simultaneously!**

This is the only conclusion available to Bob. It is simply a fact that the lightning bolts did not hit the ground at the same time. This is a statement of reality.

However, Alice was equally certain that the lightning bolts DID hit the ground at the same time. She performed exactly the same accurate measurements as Bob. For Alice, there is no doubt: it is simply a fact that the lightning bolts DID hit the ground simultaneously.

So who is right? Alice or Bob? The extraordinary truth is that they are both right — after all, they both proved their cases beyond logical doubt. As Sherlock Holmes said: "When you have eliminated the impossible, all that remains, however improbable, must be the truth." The only conclusion which is left to us is that **reality itself is different for both Alice and Bob!** In Alice's version of reality, the lightning bolts hit the ground simultaneously. In Bob's version of reality, the lightning bolts did not hit the ground simultaneously. Reality is a relative concept for the two observers. This extraordinary outcome of the experiment is called *relativity of simultaneity*.

This fundamental difference in the reality of two observers who are moving relative to each other is described by Heinz Pagels in his book *The Cosmic Code*: "Even after taking into account their relative motion and the finite speed of light they cannot agree which event 'really' took place first."

The relativity of simultaneity thought experiment which has just been described is often mistakenly presented as nothing more than an experiment in signalling, or time synchronisation. Worst of all, it is frequently presented as explaining what people "see", the observers seeing the light signals as simultaneous or not. But the experiment is not at all about seeing light signals. In fact, it is really nothing to do with light at all. What it reveals is something far more extraordinary, far more profound. It is not about what is **seen**, it is about what is **measured**. This is not about synchronisation — this is about the whole of reality. It shows that merely by moving relative to each other, two observers inhabit different realities. It is as if they inhabit two different universes.

And the events which occur in the experiment need not just be lightning bolts hitting a railway track, or lights flashing in a railway carriage. They could be a gunshot, or a star exploding, or the outbreak of war, or the assassination of a president. All aspects of reality are affected by this outcome.

The effect is imperceptible at speeds far less than the speed of light, but it is real nonetheless. If I am moving relative to you, then my reality is different to your reality.

In his book *The Emperor's New Mind*, Roger Penrose emphasizes this point by considering two people passing each other on the street at low speed. Because of the relativity of simultaneity, they will both have different conceptions of what is happening "now". While this effect is small, the effect increases with distance (as we shall discover in the next chapter). If they consider what is happening at, say, intergalactic distances then the effect is very marked. Penrose considers one person walking in the direction of the Andromeda galaxy, while one person walks in the other direction. The person walking towards Andromeda will consider his "now" moment on Andromeda to be hours or days ahead of the other person. Penrose then considers a potential invasion of Earth by aliens living on Andromeda: "Even with quite slow relative velocities, significant differences in time-ordering will occur for events at great distances. Imagine two people walking slowly past each other in the street. The events in the Andromeda galaxy (the closest large galaxy to our own Milky Way) judged by the two people to be simultaneous with the moment that they pass one another could amount to a difference of several days. For one of the people, the space fleet launched with the intent to wipe out life on the planet Earth is already on its way; while for the other, the very decision about whether or not to launch that fleet has not yet even been made!"

This effect has been called the *Andromeda Paradox*. [3]

Of course, what is happening "right now" in a galaxy 2.5 million light-years from Earth is not of particular relevance as, even if the decision to launch the space fleet is made right now, it would still take millions of years for the fleet to reach Earth. We are in no immediate danger of alien invasion.

And so we come to the end of the long journey of the last two chapters. It is a journey which started 400 years ago in renaissance Italy at the dawn of modern science when Galileo proposed the principle of addition of velocities, and introduced the modern experimental method. It is a journey which required the mathematical prowess of James Clerk Maxwell who revealed the astonishing true nature of light. And it is a journey which has ended with the genius of Einstein whose thought experiment involving the speed of light revealed the true astonishing implications for reality.

4

SPACETIME

When the Black Death struck England in the year 1665 it must have seemed like the end of the world. The bubonic plague, carried by fleas on rats, had already killed a third of Europe's population. Once you were infected, the extremities of your body turned black (hence "Black Death"), and death almost surely resulted within four days.

The closure of Cambridge University during the plague forced a 23-year-old Isaac Newton to take refuge at his home in Woolsthorpe in Lincolnshire. It was while Newton was in forced isolation that he set his mind to produce his greatest work, the *Principia*, which laid down the basis of classical mechanics and Newtonian gravity, results still very much in use to this day. Indeed, it was by using only an understanding of Newtonian mechanics and gravity that NASA put the first man on the Moon in 1969.

The following image shows Isaac Newton as painted by William Blake in 1795:

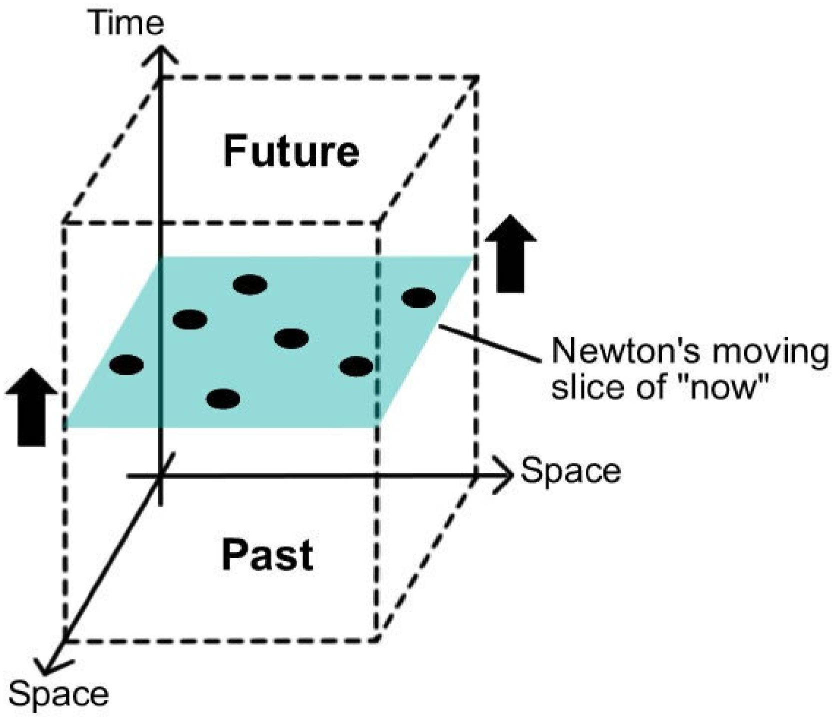
In order to arrive at his laws of motion, Newton had to employ a precise definition of time. In the *Scholium* (introduction) to the Principia, Newton stated: "Absolute, true, and mathematical time, of itself and from its own nature, flows equably without relation to anything external."

So Newton believed in *absolute time*, which always operated as a background on which all objects changed. It was as if all objects danced to a single clock which controlled the entire universe. If no objects existed in the universe, this ultimate clock would still exist counting down its absolute time.

Newton had similar views on space. Newton believed that, even if all the matter was removed from the universe, there would still be the "box" of space by which all position could be measured. This was *absolute space*.

The proposal of absolute time and space was so important and successful for Newton's theories of mechanics because it created a pre-existing framework within which objects could move, and those movements could be analysed and predicted by Newton's laws of motion.

According to absolute time, we could imagine the current state of reality in the universe as being all those events that are "real" (i.e., currently happening) at the moment we snap our fingers. That might include the boy falling off his bike down the road, and a star exploding in the Andromeda Galaxy. All of these events taken together would represent the current state of the universe. And this current "now" moment would be determined by Newton's absolute "clock of the universe" — all objects in the universe would agree on the current absolute time:



Hence, we could represent the current "now" moment as being a single slice out of all time. In the previous diagram, the current "now" moment is indicated by the shaded slice. This "now" slice moves upwards in the time direction from the past to the future (only two dimensions of space are shown on the diagram instead of the usual three dimensions).

The slice represents all the events in the current universe (the events are denoted by the black circles). Only the events included in the "now" slice are "real" — they are the only events currently happening. The events of the past have already happened and are no longer real. The events of the future have not yet happened and are therefore not yet real. As the "now" slice moves upwards along the time axis at a speed determined by Newton's absolute clock, it turns the unreal future events into real current events, and those real current events are turned into unreal past events.

However, remember back to our discussion of Einstein's thought experiment in the last chapter. We discovered that Alice and Bob — who were moving relative to each other — could not agree on whether events occurred simultaneously. In other words, they disagreed about the reality (whether or not an event was happening) of certain events. If two observers cannot agree on the reality of events, then this poses a problem for Newton's view of the universe. Newton's view was based on a slice of events which were definitely real: only the events on the slice were real. But Alice and Bob can apparently not agree which events should be on the slice — they cannot agree which events are real. An event cannot be both real and unreal at the same time. How can this problem with Newton's view of the universe be resolved?

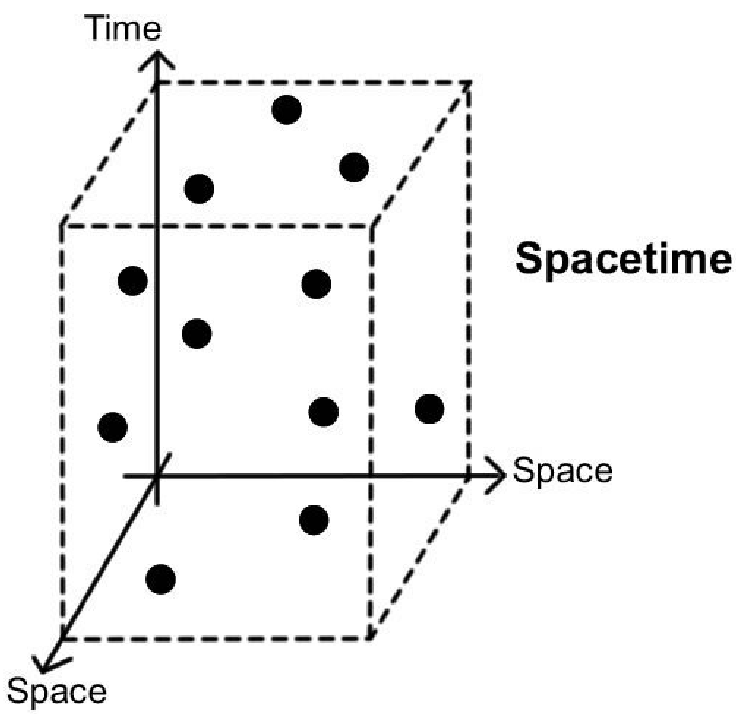
The resolution to this problem came from an unlikely source. In 1895, the English writer Herbert George Wells (better known as H.G. Wells) considered the possibility of treating time as a fourth dimension, and creating a machine which allowed the operator to move freely in that fourth dimension. After all, in our daily lives we can move in the three spatial dimensions — left/right, up/down, forward/backward — as we wish. Might it not be possible to imagine freedom to move in a fourth dimension as well? Wells's book *The Time Machine* based on this principle became the first book to popularise the concept of time travel and introduced the notion of time as a fourth dimension to the general public. We are now quite used to the idea of seeing time travel in popular fiction, so it is hard to imagine what an impact this idea must have had when presented to the public for the first time.

The following extract is H.G. Wells's description of time from *The Time Machine*. It is quite remarkable considering this was written twenty years before Einstein's great insight: [4]

"Clearly", the Time Traveller proceeded, "any real body must have extension in four dimensions: It must have length, breadth, thickness, and — duration. But we incline to overlook this fact. There are really four dimensions, three of which we call the three planes of space, and a fourth, time. There is no difference between time and any of the three dimensions of space except that our consciousness moves along it."

In truth, the idea of a fourth spatial dimension was tremendously fashionable toward the end of the 19th century. This was largely due to the efforts of popular writers such as Charles Hinton who took the latest mathematical ideas about geometry in higher-dimensional spaces and made them accessible to the general public. Another book of the time, *Flatland: A Romance of Many Dimensions* by Edwin Abbott Abbott, continued the popular trend of the time by considering how travel in extra spatial dimensions could allow liberation from conventional restraints. This idea of a liberating fourth dimension was adopted by modern artists who moved away from the restrictive one-point perspective system which portrayed the world as three-dimensional. This is especially noticeable in the perspective-free paintings of the cubists such as Pablo Picasso.

It was during this period, in this liberated environment, that Einstein's former mathematics teacher at the Zurich Polytechnic, Hermann Minkowski, took an interest in Einstein's work on relativity. Minkowski was impressed with Einstein's progress — and more than a little surprised: "It came as a tremendous surprise, for in his student days Einstein had been a lazy dog. He had never bothered about mathematics at all." Minkowski realised that an elegant explanation of Einstein's result could be provided if time was considered as a fourth dimension — just as H.G. Wells had earlier suggested. Instead of there being a single "now" slice of reality across the entire universe — as Newton had suggested — we had to consider all of time and space existing as one huge block-like structure. This is called *spacetime*.



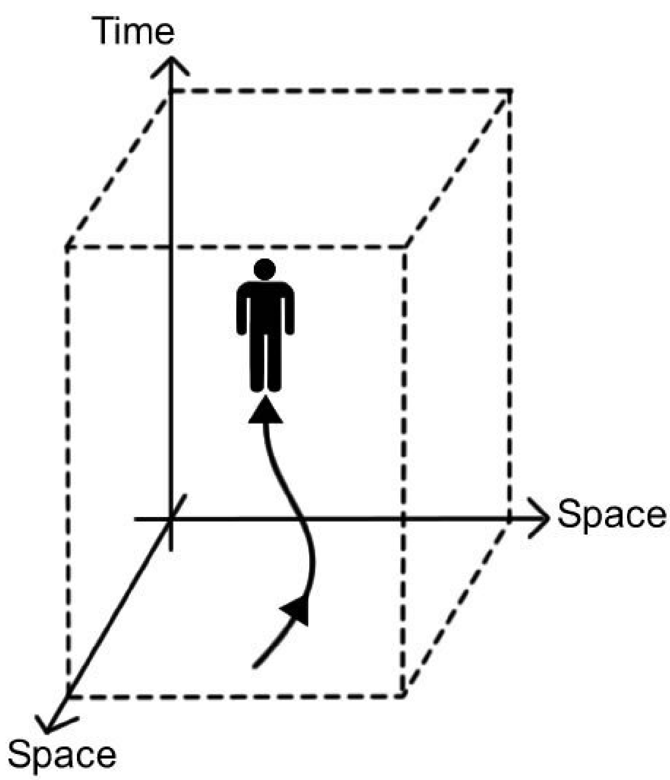
In the previous diagram, Newton's "now" slice of time has vanished, and all the events (the black circles) are now portrayed as being "real".

In principle, this presented the opportunity for time travel: if all times existed, and time was just another dimension, we could travel back to those times. This could not possibly have been the case in Newton's universe in which only the present moment was real.

It might at first appear strange to treat time as another dimension. Perhaps it helps to realise that we always give the position of events in terms of four dimensions: three dimensions of space and one of time. For example, we might arrange a meeting at the corner of two streets (providing the value of the spatial dimensions) at a certain time (providing the value of the time dimension). So events are inevitably defined in terms of four dimensions. You have been doing this all your life perhaps without realising it.

This approach, of treating space and time as dimensions of spacetime, starts to reveal the very close connection between space and time. For example, when we look at the stars we are essentially looking back in time. This is because we are looking at the stars as they were many years ago: the distances to the stars are so great that their light takes many years to reach us. Other profound connections between space and time will be considered in later chapters.

As we move around in space, we are inexorably moving forward in time. Hence, as we progress through our lives we plot a path through spacetime which is called a *world line*. In the next diagram, the world line of Bob is denoted by the curved, directional line:

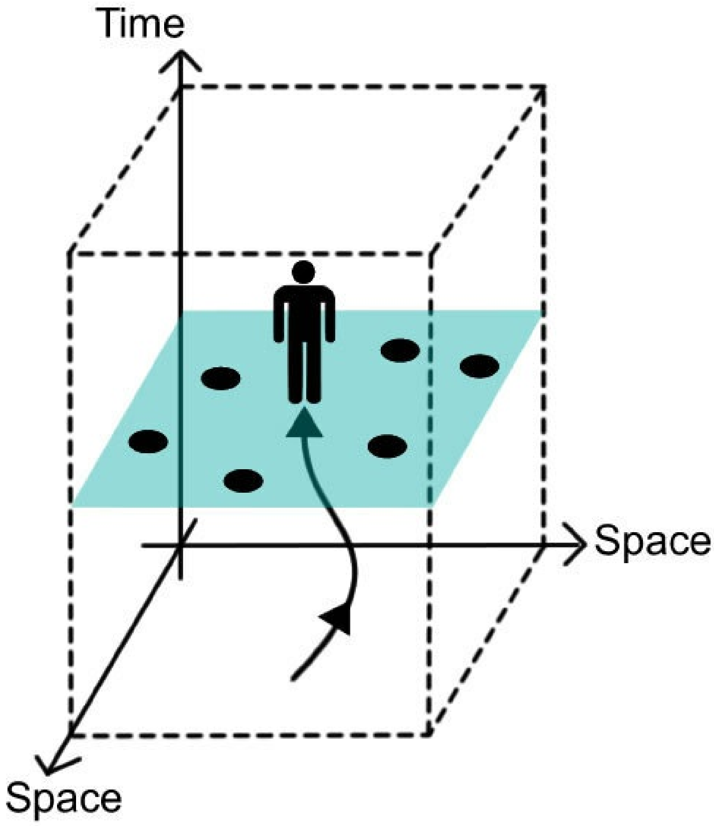


Every object moves forward in time, so every object has a world line. We might perceive an object at its single position "right now", but physics tells us that a more accurate representation of an object is as a world line through spacetime. This principle even applies to elementary particles which are portrayed in *Feynman diagrams* as lines rather than point particles.

This is one of the most important points of this book, as we shall see in later chapters. It is crucial to understanding the nature of time. Objects are not really points in space — they are truly lines in spacetime. Indeed, we ourselves exist as lines through spacetime, as if we are "stretched-out". Every atom and particle in our body exists as a line in spacetime — we just don't perceive it that way (for reasons we will discover in the next chapter).

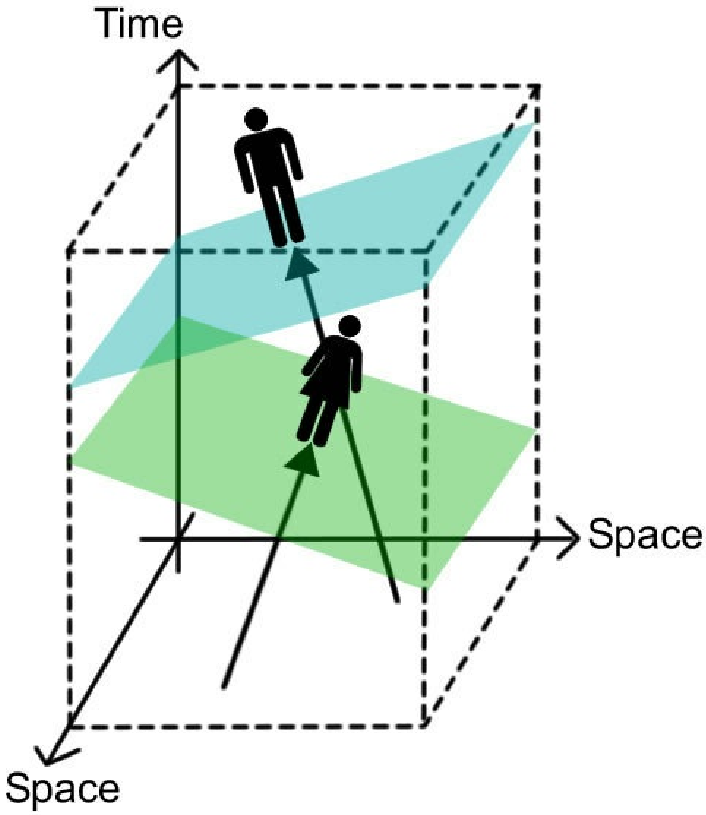
At each point of the world line of an observer, there will be an associated plane — a slice of spacetime — which represents all the events which that observer considers "real" at that particular point in time. This is called the *plane of simultaneity* of the observer.

The plane of simultaneity of Bob is shown in the following diagram by the shaded plane. Note that, at any point along Bob's world line, **the plane of simultaneity is always perpendicular to Bob's world line**. It is a cross-section of spacetime at a particular time. It represents all the events that are real to Bob when Bob clicks his fingers "right now":



This resembles Newton's "now" slice of space, but it differs in the crucial fact that **this plane is defined solely in terms of a particular observer** — unlike Newton's slice of absolute time, it does not apply to all observers. That is where Newton went wrong: he took a result which applied only to himself, and assumed it applied to all observers. He believed that the time when he "clicked his fingers" represented the same absolute time for all observers in the universe. That is not the case. Newton's model of absolute time was a flawed concept.

If we have two observers — Alice and Bob — moving relative to each other (as in the train example in the previous chapter) their world lines through spacetime will be at angles to each other. However, their individual planes of simultaneity will always be perpendicular to their world lines. This is shown in the following diagram (showing Bob and Alice's separate planes of simultaneity):



As a result of these angled planes of simultaneity, Alice and Bob do not agree as to which events are real and which are unreal (which events are currently happening, which events are yet to happen, and which events have already happened). This explains how, in a spacetime in which all events are real, Alice and Bob only experience a "slice" of real events at any moment along their world lines.

The difference in the angles of their planes of simultaneity also explains how Alice and Bob can have vastly different experiences as to which events are real at great distances. This is due to their angled planes of simultaneity diverging greatly over vast distances (this is the reason for the Andromeda Galaxy paradox discussed in the previous chapter).

So, at this stage, we have considered the basic principles of spacetime. However, some very deep questions remain:

- To say that all events in the universe are "real" all the time sounds like crazy talk! Indeed, there remains some physicists who would not agree — even though the principle has a firm basis in special relativity. Is there any additional evidence that this is the case?

- Why do human beings experience just one slice of spacetime, a "now" moment? If all events are real, why cannot we see events in the future and the past? Basically, why can't we remember the future?
- Why do we experience movement of this "now" slice from past to future?

We will be considering the first of these three questions later in this chapter. Potential answers to the other two questions will be presented in the next chapter.

The emergence of spacetime

According to absolute time and space, if all the matter was removed from the universe, there would still exist an underlying framework of absolute axes of space and time. The space and time dimensions could then be considered fundamental. However, in this chapter we have seen that Newton's model of absolute time was a flawed concept. In my previous book, it was explained how the principle that there is "nothing outside the universe" logically implies that there can be no axes of space and time — no box or clock — outside the entire universe. Absolute space and time cannot be true. So, in that case, where do the dimensions of space and time come from?

Considering time and space as being "emergent" properties is a very fashionable trend in physics at the moment. This suggests that time and space are not fundamental, pre-existing dimensions but instead "emerge" from the interactions of the objects within the universe. The popularity of this approach is largely due to a highly-influential 1997 string theory paper by Juan Maldacena. Maldacena revealed that it was possible to consider a string theory version of our universe as being a form of holographic projection of the interaction of elementary particles on the boundary of our universe. This would imply that space emerges as a result of more fundamental interactions. This result tied in very nicely with the principle of the *holographic bound* (considered in my previous book) which suggested that the maximum amount of information which can be contained in a region of space is proportional to the surface area of that region — not its volume.

Though many would disagree, Maldacena's paper remains speculative.

Most seriously, the hypothesis requires the universe to have an attractive cosmological constant (considered in my previous book) whereas the universe appears to have a repulsive cosmological constant.

So is spacetime truly emergent? Well, if spacetime is emergent, that would suggest that the dimensions of time and space are not fundamental (i.e., they do not pre-exist as underlying axes) but "emerge" as a result of the interactions of the objects within the universe. As we have discussed, the absence of absolute time and space means that the dimensions of time and space cannot pre-exist as underlying axes. Therefore, in the absence of absolute time and space, I would suggest the solution to the question is clear: time and space simply **must** be emergent, i.e., emerge from the objects **within** the universe. There are no fundamental axes **outside** the universe. Spacetime cannot, therefore, be fundamental. Logically, therefore, spacetime must emerge from the interactions of the objects within the universe.

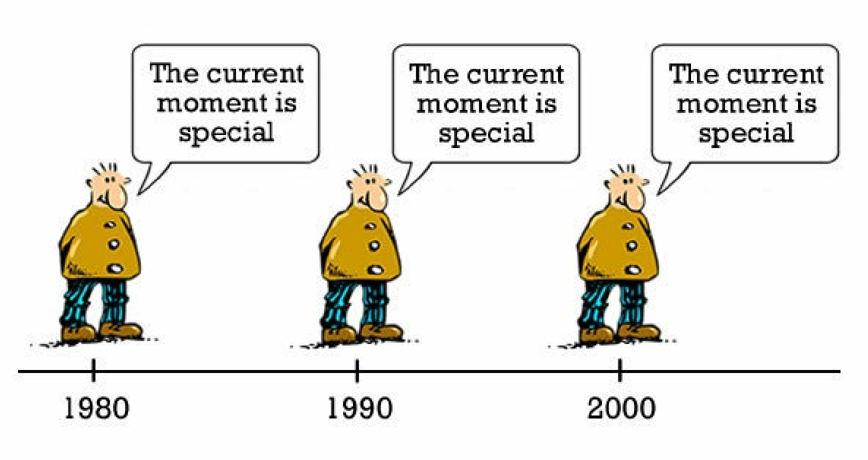
The block universe

While many readers of popular science books may be well aware of the concept of time as a fourth dimension, they perhaps still imagine the movement of a "now" moment which moves through the spacetime structure turning the future into the past. These readers are perhaps less aware of the other main conclusion of Einstein's thought experiment, which is, as we have seen, that all events — and therefore all times — are equally real. There is therefore no special "now" moment. As Brian Greene notes in his book *The Fabric of the Cosmos*: "A less than widely appreciated implication of Einstein's work is that special relativity really treats all times equally."

What this implies is truly quite staggering. It suggests that all times are real, every point of our life has the same level of reality. The time when you fell off a bike when you were a child is just as real as the current moment. And the moment of your birth is just as real as the moment of your death. In fact, all times really exist — the whole block of spacetime from the start of the universe to the end of the universe exists as one unchanging block. This is called the *block universe* model.

I considered the block universe model in depth in my first book, and I do not want to repeat myself, but it has to be mentioned here as it appears to be one of the few vitally-important genuine insights we can obtain about the nature of time.

But, you might argue, you feel as though there is a moving "now" moment. Surely time is moving? Surely "now" is a special moment? Well, yes, you do feel a moving "now", but that is precisely how you have felt at every other moment of your life:



Time is not actually moving — there are just multiple copies of you stretched-out through spacetime. The moving "now" slice is just an illusion of human perception (which we will consider in the next chapter).

In the last section I asked the question as to whether any additional evidence existed to back up the conclusion of special relativity that all times exist. And, indeed, it seems basic logic can prove the existence of the block universe model. This is because we can easily recognise that the concept of a moving "now" is logically inconsistent. Put simply, it makes no sense to ask the question: "How fast does time flow?" When we generally talk of the speed of moving objects, we give an answer which represents the change in position of the object **with respect to time**. However, when we talk about a moving "now" moment we are talking about the movement of time itself, and it makes no sense to talk about the speed at which time itself moves **with respect to itself**. To say that time moves at a rate of one second per second is absurd. The only logical conclusion is that time does not flow — all times are real.

So not only does special relativity tell us that we live in a block universe, basic logic tells us that it is undoubtedly the case.

The block universe in literature

If "killing your own grandfather" paradoxes are a favourite cliche of schlocky science fiction, then a more accurate and sophisticated model of time incorporating the block universe model can be found in more intelligent sci-fi literature. These thoughtful portrayals of the block universe present time as an unchanging, pre-existing block structure which can be accessed by beings possessing extraordinary powers.

As an example, Alan Moore's magnificent *Watchmen* graphic novel includes the character known as Doctor Manhattan. Doctor Manhattan is created when the physicist Dr. Jonathan Osterman is accidentally disintegrated when he is trapped in an Intrinsic Field Subtractor (the "intrinsic field" apparently being the force which holds all matter together). His body slowly reassembles itself into something greater: a blue-skinned super-being with control over matter at a subatomic level. Osterman is pressed into service by the U.S. government which gives him the name Doctor Manhattan (named after the Manhattan Project).

Doctor Manhattan is an almost omnipotent figure, and possesses the ability to perceive all times at once. It is as if Doctor Manhattan exists "outside the universe" and is in possession of a godlike overview of the entire block universe structure. As Doctor Manhattan explains: "There is no future. There is no past. Do you see? Time is simultaneous, an intricately structured jewel that humans insist on viewing one edge at a time, when the whole design is visible in every facet."

While Doctor Manhattan's character might appear rather distant and otherworldly, his realisation and description of the block universe is instructive and carries a message for all of us. All times exist equally — this is what science and logic tells us. Try and see the "whole design visible in every facet" rather than concentrating on "viewing one edge at a time" (i.e., the "now" moment). As you can tell, Alan Moore writes the Doctor Manhattan passages in *Watchmen* quite beautifully and poetically.

As another example of the use of the block universe model in quality literature, in 1969 the science fiction author Kurt Vonnegut released his most famous novel, *Slaughterhouse-Five*, now considered one of the finest novels of the twentieth century. *Slaughterhouse-Five* is the story of Billy Pilgrim who is abducted by aliens from the planet Tralfamadore. The aliens are friendly, and can see in four dimensions (i.e., including the time dimension). Hence, the aliens possess a similar godlike overview of spacetime to that enjoyed by Doctor Manhattan. The Tralfamadorians see it as their duty to teach Earthlings about the true nature of time. As Billy Pilgrim recalls: "The most important thing I learned on Tralfamadore was that when a person dies he only **appears** to die. He is still very much alive in the past, so it is very silly for people to cry at his funeral. All moments, past, present, and future, always have existed, always will exist. The Tralfamadorians can look at all the different moments just the way we can look at a stretch of the Rocky Mountains, for instance. They can see how permanent all the moments are, and they can look at any moment that interests them. It is just an illusion we have here on Earth that one moment follows another one, like beads on a string, and that once a moment is gone it is gone forever."

The Tralfamadorians were able to see the true nature of objects (and people) stretched-out as world lines in spacetime, a continued existence from baby to old age: "Tralfamadorians don't see human beings as two-legged creatures, either. They see them as great millipedes — with babies' legs at one end and old people's legs at the other."

As I said earlier in the chapter, this is an absolutely crucial point in this book: we do not exist as points in time, we are "stretched-out" as world lines from birth to death — like a millipede. This is the reality of our existence in spacetime. We will be seeing in later chapters why this is so important.

How to live in a block universe

The implications of the block universe model really seem quite staggering. It states that you exist at all times: the past and the future are just as real as the current moment. So the question arises: how should you live your life in a block universe? Does an understanding of the block universe give you an advantage over other people (who don't read quality popular science books)? Should you behave differently?

Firstly, I think you should find living in a block universe reassuring. Life can feel as though we are on an inexorable conveyor belt, from birth to death. However, once you realise that the motion of time is just an illusion, and that all times are equally real, you realise that the conveyor belt of time is just an illusion of your mind. You can step off the conveyor belt. You can appreciate that you are alive at all times, that you exist at all moments in spacetime. You should feel relieved and empowered.

And death should hold less fear. True, the block universe model implies that the circumstance and time of your death has already been decided, but it also states that the moment of your death has no more importance than any other time of your life. You will remain alive at other times — for eternity. This is no crank theory — this is an underappreciated implication of orthodox physics.

Michael Lockwood considers this principle in his book *The Labyrinth of Time*. He tries to console those who may have lost loved ones: "Einstein evidently believed that the spacetime view, when taken fully to heart, can provide comfort to the bereaved. A person who is not living **now**, but did or will live at other times, exists in just as substantial a sense as someone who does not live **here**, but only in some other **place**. Einstein is urging us to regard those living in times past, like those living in foreign parts, as equally **out there** in spacetime, enjoying the same flesh-and-blood existence as ourselves."

If all times are equally real, then this implies that those happy moments you have in your past are just as real as the current moment. That should be reassuring. But it also raises the intriguing question of whether it might be possible to raise your level of awareness of past events to the same level of awareness you have of the current moment, to relive the past. I believe this is possible — though not in the obvious manner. It does not appear possible to raise your awareness of the past, so, instead, lower the importance you give to the current moment. Downgrade the "now". Realise that the current moment is no more real than the past moments in your life. We are inundated with advertising slogans which try to persuade us to increase the importance we assign to the current moment: "Live for the day!" "Buy now, pay later!" Instead, resist these temptations. Do not "live for the day", live for all times. Realise that you exist at all times. Save for the future. Appreciate the past as being as real as the current moment.

When it comes to living in a block universe, I think we could all learn from Doctor Manhattan and Billy Pilgrim.

5

THE ARROW OF TIME

As we have just discussed, the block universe model suggests that all of spacetime exists as an unchanging block, with no special "now" moment. But, if that is the case, then why do we feel as though there **is** a "now" moment, which gives the current moment priority over the past and the future? In fact, not only do we feel a "now" moment, but we feel movement of that now moment from what we call "the past" to what we call "the future". So this seems to indicate a directionality in time. What is the origin of this directionality?

This directionality — this so-called *arrow of time* — is not just limited to our internal perception of the motion of time. External physical processes also exhibit a directionality in time. For example, we might see an egg breaking (in the forward time direction), but we never see a broken egg reforming itself. These considerations might seem trivial, but they pose serious questions for physics. This is because the laws of physics — certainly Newtonian mechanics — are *time symmetrical*, i.e., they work the same in the backward time direction as they do in the forward time direction. Consider, for example, shooting a movie of two balls colliding, with the first ball stopping and the second ball moving off at speed. If you played the movie backwards, the events would still make sense according to the laws of physics. This time, though, the second ball would come in reverse, strike the stationary first ball, and the first ball would then move off in reverse. Everything would happen perfectly in reverse, and it would look as though it was happening in the forward time direction.

So if the laws of physics are time-symmetrical, why do so many processes exhibit an arrow of time in the forward time direction?

Let us first consider the psychological aspects of this question.

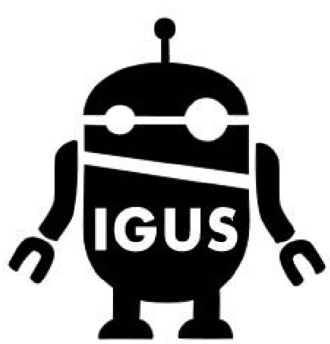
The feeling of "now"

In the discussion of the block universe in the previous chapter, it was described how all times are equally real and there is, therefore, no special "now" pointer which moves through time at a certain speed and determines the current moment. Even though we derived this result in a logical manner, there still lingers a considerable amount of resistance to this model. For example, Lee Smolin recently devoted an entire book, entitled *Time Reborn*, in an attempt to refute the block universe. The main reason I feel that many people have such a problem accepting the block universe model is not through any rational scientific basis, but because they instinctively feel a special "now" moment. This feeling is so utterly entrenched into our lives and psyche that it is an incredibly hard habit to break.

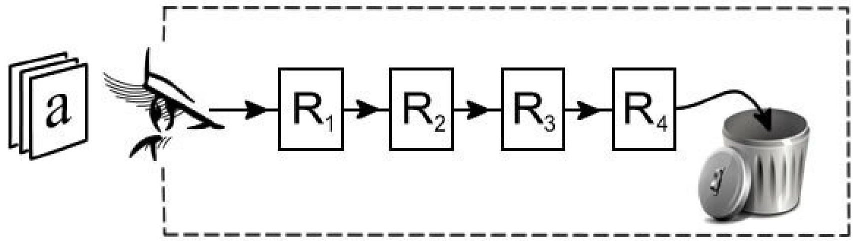
However, there is no place in physics for feelings. Physics is based on measurement and cold equations. Is it possible to produce any hard numerical measurement of this supposed "moving now"? No, of course it isn't — for the reason presented in the previous chapter: it simply makes no logical sense to try and measure how fast time flows. People might say with complete certainty that they feel the movement of a "now" pointer, but you will note that they never say precisely how fast it moves!

So why do we feel this motion of time, this so-called *psychological arrow of time*? The physicist James Hartle believes the answer does not lie with physics at all, but instead purely lies in human biology. Hartle has considered how the human mind processes information, specifically considering the mechanism of human memory.[5] Hartle realised that a computer model of human memory could be constructed. The resultant model applied not only to humans, but also to processing units with computer memory such as robots. These generalised robots were given the name IGUS (Information Gathering and Utilizing Systems).

Here is a cute example of an IGUS. We'll be seeing him again later:



Hartle realised that the memory of an IGUS could be modelled as a series of computer registers:



In the preceding diagram, the IGUS mechanism is contained within the dashed rectangle. The ever-changing information about the outside world is denoted by a pack of cards (shown on the left of the diagram), with the top card changing with time. The latest information about the external world — the top card on the pack — is captured by the electronic eye of the IGUS. That information is stored in its first memory register, R_1. As time passes, that information in register R_1 gets shuffled back further in memory into register R_2, and the new latest information is captured into R_1. So as each new piece of information comes in, the information which is already held in memory gets shuffled further and further back in the memory registers. Eventually, the information is shuffled out of the bank of registers completely, into the waste bin, and the information is lost ("forgotten").

Now, the important point here is that our brains **could** assign equal priority to each of these memory registers. It could say the contents of R_2 are just as important as the contents of R_1, or it could even say that the contents of R_4 are **more** important than the contents of R_1! But that is not what happens in practice. In fact, any creature whose brain worked on the principle that the contents of R_4 were more important than the contents of R_1 would be quickly killed off through evolution. This is because such a creature would be perpetually living in the past. It would be acting according to the state of the universe a few seconds ago. Hartle gives the example of a frog trying to use its sticky tongue to catch a fly on a leaf. If the frog is operating based on the contents of R_4 then it will be considering the state of the universe a few seconds ago. By the time the frog sticks its tongue out to catch the fly, the fly will have long since flown.

So, instead, our brains assign greater importance to the contents of memory register R_1. Evolution has shown that this is the most efficient way to catch flies. The brain still retains the contents of memory registers R_2 to R_4, but it makes their contents appear "foggy" to us, to make it clear that these contents should not be regarded as being as important as the contents of R_1. Hence, we call the contents of R_2 to R_4 "memories".

The brain retains the contents of R_2 to R_4, and makes their contents available to us, because memories are useful for evolutionary purposes. But the brain emphasizes the importance of the contents of R_1. If the brain really wanted to, it could raise the importance of the contents of R_2 to R_4 to the same level of importance as the contents of R_1. Our memories would then be perceived as being just as real as the current moment. We would then be truly like the Tralfamadorians from *Slaughterhouse-Five*, viewing all past moments in time as being as equally real as the current moment.

How extraordinary! And perfectly legitimate within the laws of physics.

So the feeling of a special "now" moment is purely a construction of biology. Do not be fooled — this has nothing to do with physics. Physics states that all times are equally real.

But it is interesting that this is how an IGUS (such as ourselves) can generate a "now" moment from the static block universe structure: it is purely an illusion of human perception.

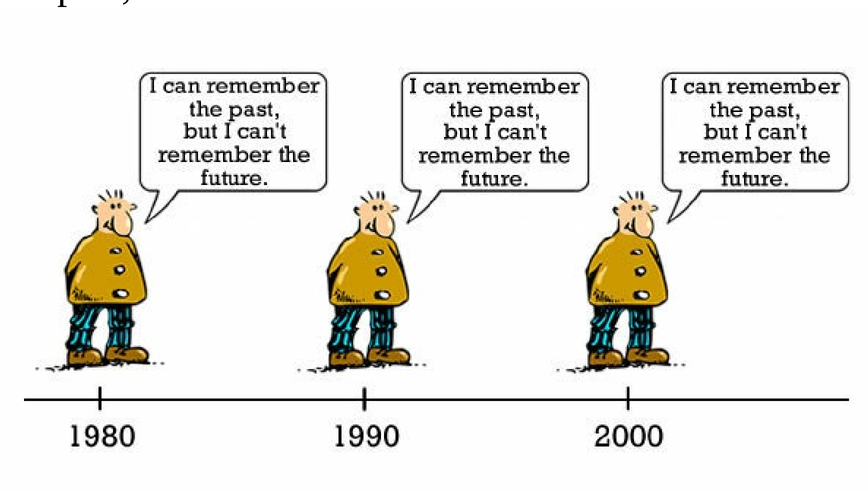
This is why we do not see objects as continuous world lines in spacetime (like the Tralfamadorians). We only ever regard the true state of the universe as being the information which resides in the latest register R_1. So we only ever see a cross-section of spacetime at one point "right now". In this respect, human perception fools us — it does not provide us with an accurate view of reality. As I stressed in the previous chapter, it is vital to understand that the correct view of reality is that objects exist as stretched-out world lines in spacetime.

So, if this book is a detective story trying to solve the true nature of time, this tells us we cannot trust the testimony of our witnesses.

The moving "now"

So, given that the universe has a static block universe structure, why does an IGUS feel the movement of this "now" moment? In other words, what is the reason for the psychological arrow of time? To answer this, you will notice that there is a clear directionality in the way the captured information passes through the IGUS memory registers: the information always passes from left to right, from register R_1 to register R_4. It is always one-way traffic. This directionality results in an asymmetry in our memory: we can remember the past, but we cannot remember the future. This asymmetry is the reason we feel movement: we feel the past is somewhere "we have been" and the future is somewhere "we are going".

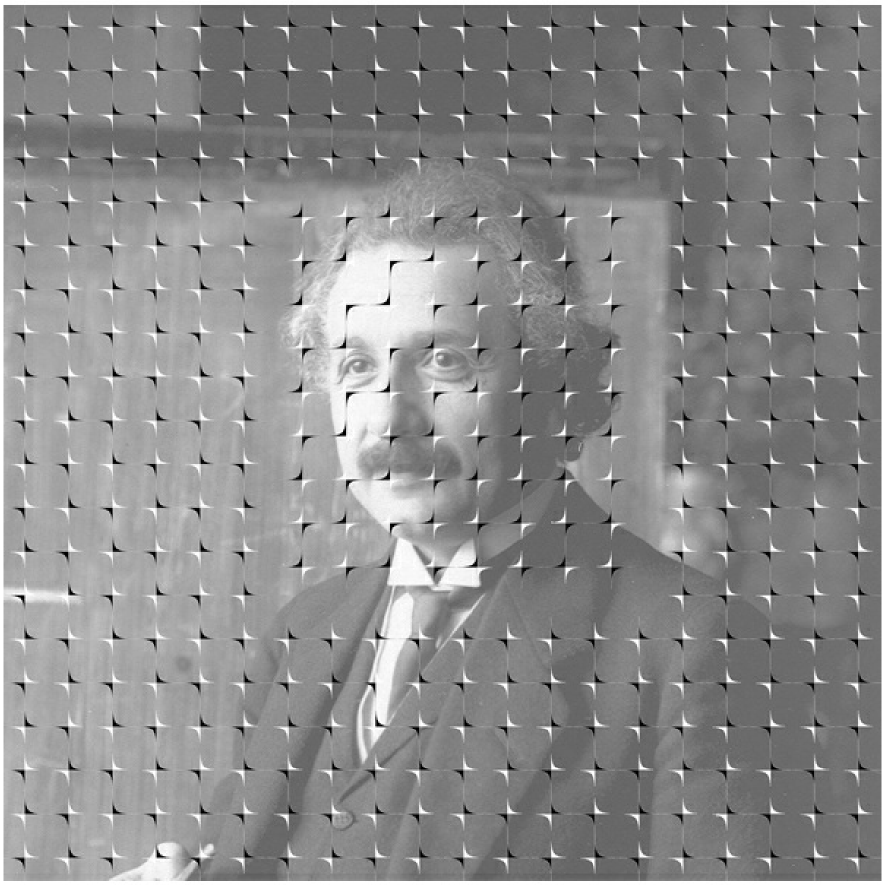
If we return to consider our cartoon character who lives in the block universe, we find that at each point in his existence he is able to remember his past, but is unable to remember his future:



So this asymmetry results in our cartoon character feeling a form of motion in time, directed from the past to the future. However, due to his existence at all times in the block universe model, there is no actual motion at all. The universe is a static block of spacetime.

This sensation of a "moving now" explains how humans can observe motion in an otherwise completely static block universe structure. The storage registers of our memory hold past states of our environment. As the data shuffles through our memory registers, we can compare the past state of the environment with the current observed state. Any change in the state of our environment would be interpreted as "motion" of the objects in the environment. So, even in a completely static block universe structure, human beings can obtain an impression of motion. There are probably good evolutionary reasons why we have developed a memory which works in this manner: it is undoubtedly a huge evolutionary advantage to be able to detect objects in motion. This is the reason why a tiger creeps slowly through the African Savannah when it is tracking its prey — it is because the visual system of a gazelle is attuned to detecting motion.

So what science and logic seems to be telling us is that the sensation we have of movement through time is nothing more than an illusion generated by our brains! You might find this very hard to accept, after all, surely our brains never fool us into believing that something which is static appears to be moving? Well, if you think that is the case, consider the following optical illusion:



You should be able to see motion in the image (Einstein's head moves).

So the brain clearly has a tendency to give an impression of movement in a situation which is completely static. There is much in physics which is counter-intuitive, and placing too much credence on the evidence of our own eyes — and our "feelings" — can hinder our progress to uncover the truth. We have to be careful only to analyse the available data, and not impose our own preconceptions on the result.

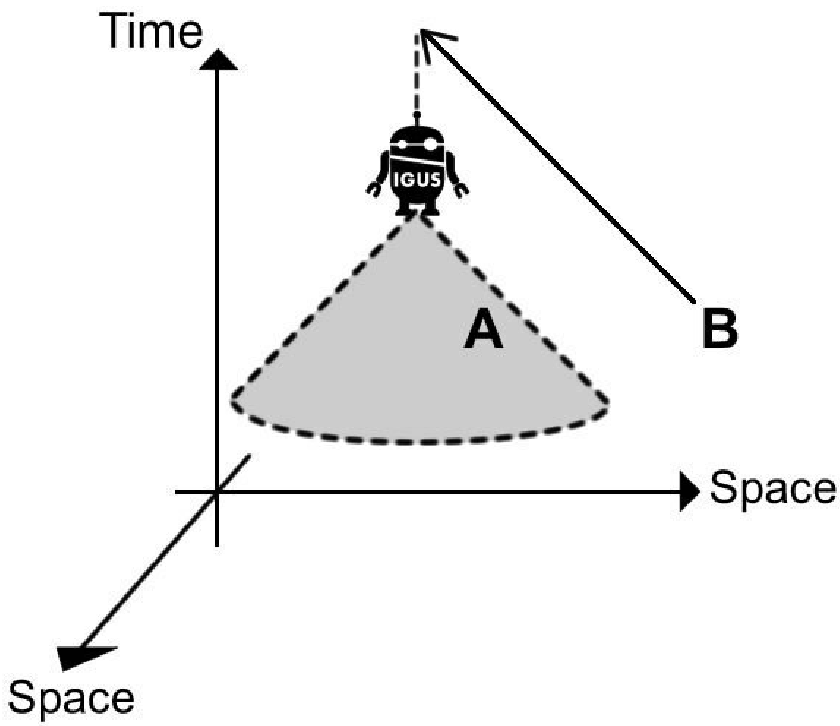
Why can't we remember the future?

So, as has just been described, the underlying reason as to why we experience a moving "now" moment is due to an asymmetry in our memory: we can remember the past, but we cannot remember the future. This asymmetry generates a feeling of movement: we feel the past is somewhere "we have been" and the future is somewhere "we are going". So, as we continue our analysis, it seems essential to try to understand the root cause of this asymmetry, which is: why can't we remember the future?

It might be imagined that the answer to this question is irretrievably locked in the maze of neurons in our head. The answer would appear to be a very personal matter, the result of individual brain chemistry. However, once again James Hartle's IGUS model casts light on the mystery by considering the laws of physics.

Any event which occurs in spacetime has a *light cone* associated with it. In this example, we will be considering the light cone for past events (there is also a corresponding light cone for future events). The past light cone contains every event which might possibly cause a particular event. Basically, the past event must have occurred close enough to the future event so that light had enough time to reach the future event. This allows the past event to have been the cause of the future event.

The past light cone for an IGUS is shown in grey in the following diagram:



The IGUS lies at the apex of the cone. Two events, A and B, are shown on the diagram. As can be seen, event A lies within the past light cone of the IGUS. This means that light (and, hence, information) from the event has had sufficient time to reach the IGUS. As far as the IGUS is concerned, event A is an event in its past. The IGUS possesses information about the event.

However, it can be seen that event B lies outside the past light cone of the IGUS. This means that light (and information) from event B has not had enough time to reach the IGUS. Light from event B will only reach the IGUS after some additional time has passed (see how the arrow from event B intersects the IGUS world line some time in the future of the IGUS). Hence, event B can be considered to lie in the future of the IGUS. As far as the IGUS is concerned, event B is a future event.

So why is the IGUS unable to remember the future? Well, James Hartle makes the rather obvious and brilliant insight that "One reason the robot doesn't remember the future is that it receives no information about it." There has been enough time for information from event A (the past event) to reach the IGUS, so information about that event has been captured into the memory registers of the IGUS. However, there has not been enough time for information from event B (the future event) to reach the IGUS. Information from event B has therefore not been stored in the IGUS memory registers. Whereas the IGUS can remember event A (the past event), it is unable to remember event B (the future event). **The IGUS is therefore unable to remember the future.**

So, far from being a matter of personal brain physiology, we discover that the reason we cannot remember the future is purely due to the physics of the external universe: we cannot remember the future because we have received no information from the future. And this is an important message: the boundary between the internal processes of our minds and the external processes of the universe is a very blurred boundary. Human beings are not the isolated entities we often imagine we are.

The thermodynamic arrow of time

This behaviour of light which has just been described — always travelling from the past to the future — appears to represent another directionality in time, another arrow of time. This behaviour is actually called the *radiative arrow of time*. There are in fact many different "arrows of time". For example:

- We have already discussed the *psychological arrow of time*, and examined why we feel the motion of time from the past into the future.
- We have also considered the *radiative arrow of time*, and seen its importance in generating the psychological arrow of time.
- There is also a *quantum mechanical arrow of time*. If you read my first book, *Hidden In Plain Sight*, you will know that before we measure a property of a particle, the particle behaves as though it has **all** possible property values. However, when we measure the actual property of the particle, there is an apparent "quantum jump" to a particular value. This quantum jump always happens in the forward time direction.
- Another arrow of time is the *thermodynamic arrow of time*. Heat will always flow from a warm body to a cold body.

Interestingly, these seemingly unrelated arrows of time all have several features in common. Firstly, all of these arrows of time represent an **irreversible** process. In the psychological arrow of time, we feel movement in the forward time direction, but we cannot reverse our motion to feel as if we are travelling in the reverse time direction. In the radiative arrow of time, light travels in all directions away from a source and scatters randomly off objects, but scattered light is never observed converging back towards a source. In the quantum mechanical arrow of time, quantum jumps happen in the forward time direction, but can never happen in the reverse time direction. And in the thermodynamic arrow of time, heat flows from hot to cold, but the reverse process is never observed (the heat of a body at constant temperature does not spontaneously move to one corner of the body for no reason).

The second similarity between all of these arrows of time is that all of these irreversible processes happen in the same forward time direction.

So, these similarities lead us to quite an intriguing thought: might it be possible that all of these arrows of time have the same underlying cause? That would explain the similarities. In fact, it is indeed now the case that physicists believe that all of these arrows of time have the same cause. And, perhaps surprisingly, the underlying cause is believed to be the principle behind the thermodynamic arrow of time.

Thermodynamics is the theory of heat and motion ("thermo" + "dynamics"). Heat always flows from a hot substance to a cold substance, and this motion can be tapped — just like the motion of water can be tapped to drive a water wheel. This motion provides the energy for steam engines, for example. And it was during the era of the steam engine that the science of thermodynamics first appeared.

So why does heat always flow from hot to cold? Well, what is really happening is that the heat is moving from being concentrated to being dispersed. If you have two objects — one hot and one cold — and you connect them to form a single system, the eventual result will be a single system at the same temperature. In other words, heat has flowed from the hot object to the cold object in order to equalise the temperature of the two objects.

In 1865, the German physicist Rudolf Clausius observed this flow of heat and tried to express the motion in mathematical terms. He arrived at the following inequality:

$$\frac{dS}{dt} > 0$$

This inequality says that the rate of change of a measurable property, *S*, in a thermodynamic process is always greater than 0. In other words, the value of *S* always increases. Clausius gave a name to this peculiar property, *S*: he called it *entropy*. The inequality therefore states that the value of entropy always increases with time. This forms the *second law of thermodynamics*.

So at last this equation gives us a clear mathematical formulation for an arrow of time: it is the arrow of time captured in an equation! We are not talking about a subjective psychological arrow of time here — this is a measurable physical quantity which always increases with time.

However, despite the success of the laws of thermodynamics, in the middle of the 19^{th} century no one quite understood what entropy was, or why it should always increase. In order to make sense of entropy, another conceptual leap was required. That leap was the acceptance of the existence of atoms.

At the time, the existence of atoms was still in doubt. However, if we consider objects to be made of atoms then we can realise that heat is simply the random motion of atoms. The spread of heat through an object is then the spreading of this random atomic motion through the object.

The great Austrian physicist, Ludwig Boltzmann, realised that if we consider our world as being made up of atoms, we could obtain an understanding of entropy, and why it must always increase. If we have a system made up of atoms, we could consider those atoms to be in either an ordered state, or a totally random disordered state (this is not something you could do if you considered objects to be a continuous solid). Entropy could then be regarded as the amount of disorder, or randomness, of the atoms in a system. For example, if heat energy was in a very ordered state, confined to a few atoms vibrating at the edge of an object, that heat energy would tend to spread through the whole object. In other words, order would be lost: disorder increases. In the final state, all the atoms in the object would be randomly vibrating, so the whole object would be the same temperature, with no trace left of that initial ordered state. This would be a state of maximum disorder, maximum entropy.

Boltzmann produced an equation which allowed you to assign a value to entropy — the amount of disorder:

$$S = k \log W$$

This is Boltzmann's famous equation, engraved on his tombstone in Vienna.

The value for the *W* term in the equation is produced by examining a system and calculating how disordered is its current state. For example, a pack of 52 playing cards could be randomly shuffled into a huge number of different sequences. If, however, you found the pack perfectly ordered in sequence, ace to king in all four suits, then this would represent a very special, ordered state. And that sequence would represent a very small proportion of all the possible sequences. Hence, the *W* value would be very small for that perfect sequence — giving a very small entropy value in Boltzmann's formula.

In the case of maximum entropy, the pack would be completely disordered, and the *W* value would then be equal to the total number of possible sequences of the cards: a huge number.

So, by considering Boltzmann's formula, why should entropy always increase? Well, a system will have many more disordered states rather than ordered states. Consider the pack of playing cards again. If the entire pack of cards was ordered in sequence, that represents a very special, ordered state with low entropy. However, a pack of cards has many more disordered states. If the ordered pack was then randomly shuffled, it would most likely move to one of the disordered states — it would be highly unlikely to move to the ordered state. Because systems have many more disordered states than ordered states, a system which changes state randomly will tend to move to a more disordered state.

So this explains the thermodynamic arrow of time. It explains why an area of heat (random atomic motion) spreads, and never shrinks. And this principle — that disorder will spread — applies to all dynamic systems. It explains why we might see a china cup breaking into hundreds of disordered pieces, but we never see broken china pieces spontaneously reforming to form a cup. To quote Jim Al-Khalili from his BBC TV programme *Order*

And Disorder: "Boltzmann's equation contains within it the mortality of everything, from a china jug, to a human life, to the universe itself."

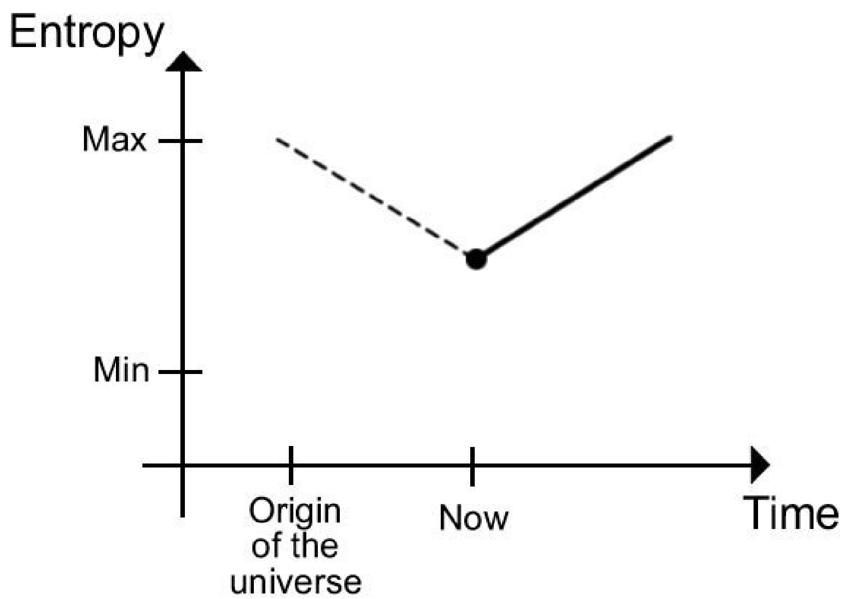
Loschmidt's paradox

However, we are not out of the woods yet. There is still a problem with our proposed solution to the thermodynamic arrow of time. If you remember back to the start of this chapter, it was described how the laws of physics are fundamentally time-symmetrical, i.e., they work the same in the backward time direction as in the forward time direction. So the laws of physics do not favour any direction of time — forward or backward. And, surprisingly, when we consider Boltzmann's solution we also find that no direction of time is preferred. Boltzmann's solution just says "a system that changes randomly will tend to become more disordered." But, if you consider the backward time direction, systems change just as much in the backward time direction as they do in the forward time direction. Again, no particular direction of time is preferred.

So change of entropy is fundamentally time-symmetrical. I do not think this point is widely appreciated.

Hence, it appears that we should find entropy increasing in the reverse time direction just as much as we observe entropy increasing in the forward time direction. This argument is called *Loschmidt's paradox*, named after Josef Loschmidt who was Ludwig Boltzmann's teacher.

Now let us apply Loschmidt's paradox to the whole universe. As change of entropy is fundamentally time symmetrical then that suggests that we should see entropy increasing as we look back in time from our current moment. This increase of entropy into the past is indicated by the dashed line in the following diagram:

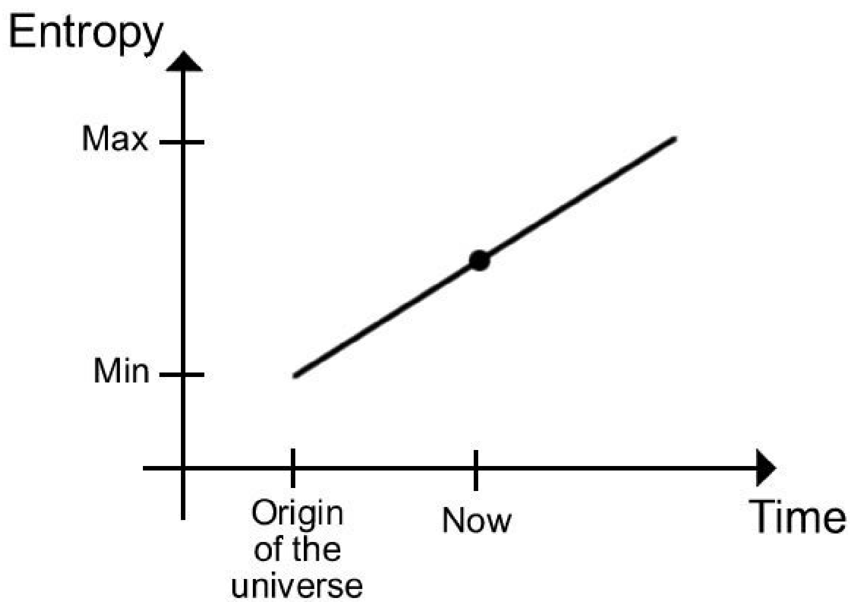


Note the symmetry of the "V" shape. This is due to change of entropy being fundamentally time-symmetrical.

However, this is, of course, not what we observe. Objects tend to have lower entropy in the past — not higher entropy as this diagram suggests. Consider a rusting old car, which will continue rusting into the future (i.e., increasing its entropy). However, if we look into the past we will find a sparkly new car with lower entropy. So as we look into the past we find objects with lower entropy — not higher entropy.

So why do objects have lower entropy in the past? This is due to the very special state of the universe when it was created. The universe was created with incredibly low entropy, and all the objects contained within the universe have been essentially falling apart ever since!

The following diagram shows the incorrect dashed line removed, and replaced with a line revealing the extremely low entropy at the origin of the universe. It shows the entropy value steadily increasing throughout the lifetime of the universe:



So why was entropy so incredibly low at the origin of the universe? We are not yet certain as to why that was the case, but a possible solution may be found from an unlikely source …

Entropy and biology

There was a story in the news recently of a remarkable young girl from Montana called Gabby Williams. Gabby is 8 years old, but resembles a baby. The rate at which she is ageing has slowed to a snail's pace. Gabby is not the only person with this condition: there is a 29-year-old man from Florida who has the body of a 10-year-old, and a 31-year-old Brazilian woman has the body of a 2-year-old. This condition is so rare that doctors do not even have a name for it.

These remarkable people have come to the attention of a medical researcher called Dr. Richard Walker, now retired but currently based in St. Petersburg. Dr. Walker believes these people could hold the secret to immortality within their genetic structure. He refers to the process of ageing as "developmental inertia" (which is a very interesting term with particular relevance to physics — as we shall see later). Walker believes that it might be possible to identify an "off switch" for developmental inertia, and thus avoid the ageing process altogether.

However, if Dr. Walker is to achieve immortality he will have to overcome a fundamental biological limit. It appears that a human cell can only divide about 50 times. This limit on cell division is called the *Hayflick limit*. At the end of each strand of DNA there is a repeated sequence of molecules which act to stop the DNA from fraying — much like the piece of plastic on the end of shoelaces. With each cell division, this piece of "plastic" DNA shortens. So this behaves like a molecular clock, with each cell having a self destruct mechanism which is designed to limit life span. When the clock expires, the cell undergoes programmed cell death. The Hayflick limit is different for other animals: if you are lucky enough to be a Galapagos turtle your cells can divide about 110 times.

If Dr. Walker is right, it might be possible to circumvent the Hayflick limit. He realises that this would not necessarily lead to eternal life: he admits that a serious accident or disease is still going to kill you. However, if you can avoid getting run over by a car, or contracting ebola, it would appear that Dr. Walker does indeed suggest that living forever is a real medical possibility.

However, as Dr. Walker appears to have a preference for adopting physics terms in his research, perhaps someone might introduce him to the concept of what we might call "developmental entropy". Fundamentally, a human being represents a physical system which is inevitably prone to increasing entropy. In this respect, a human is no different from a rusting car, or a breaking egg. A human has to follow the unidirectional arrow of time from cradle to grave, as there is simply no way of circumventing the second law of thermodynamics. And the second law of thermodynamics certainly applies to cell division.

A mature human is made up of several trillion cells. When these cells become old and require replacing (e.g., skin cells are constantly being shed), new cells can be produced by *mitosis*. Mitosis is cell division: one old cell can produce two identical daughter cells. It is essential that the DNA in the original cell is precisely copied to the two new daughter cells. However, during this copying of the DNA, it is possible for errors to creep into the copied DNA, and these errors can accumulate over time. Imagine taking a photocopy of an image, and then photocopying that photocopy, and then repeating the process many times. Eventually, you will find many errors creeping into your photocopied image so that it no longer resembles the original image.

Unfortunately, these accumulated errors in the DNA can produce cancerous cells. Cancer is generally described as the result of accumulated DNA errors, and the role of carcinogens (cancer-causing chemicals) is often listed as the main cause. However, we can clearly also view the accumulation of DNA errors as an inevitable result of increasing entropy. This explains why cancer is predominantly a disease of old age. As the cancer researcher Robert Weinberg explains: "Cancer is an inevitability the moment you create complex multicellular organisms and give the individual cells the license to proliferate. It is simply a consequence of increasing entropy, increasing disorder. If we lived long enough, sooner or later we would all get cancer." [6]

Interestingly, though, it appears that a certain strain of cancer can avoid the seeming inevitability of entropic damage. In Baltimore in 1951 some cancerous cells were removed (without permission) from a tumour in Henrietta Lacks. Scientists were stunned to discover that these cells were apparently immortal. Whereas normal human cells could survive for only a few days, these cells (known as HeLa, after the initial letters of Henrietta Lacks' name) could survive and replicate without limit. These immortal cells have been grown and distributed around the world, and have been used in countless experiments. Henrietta Lacks left a tremendous legacy to the medical world, and has undoubtedly saved many lives. In many laboratories around the world, Henrietta Lacks continues to live.

So how can HeLa cells be immortal? How can they avoid inevitable damage due to increasing entropy and damaged DNA during replication? Well, the truth is that even HeLa cells are subject to increasing entropy. Cell division in HeLa continues to damage the DNA and produce mutations, which results in new strains. It appears that absolutely nothing can escape the degradation due to the second law of thermodynamics.

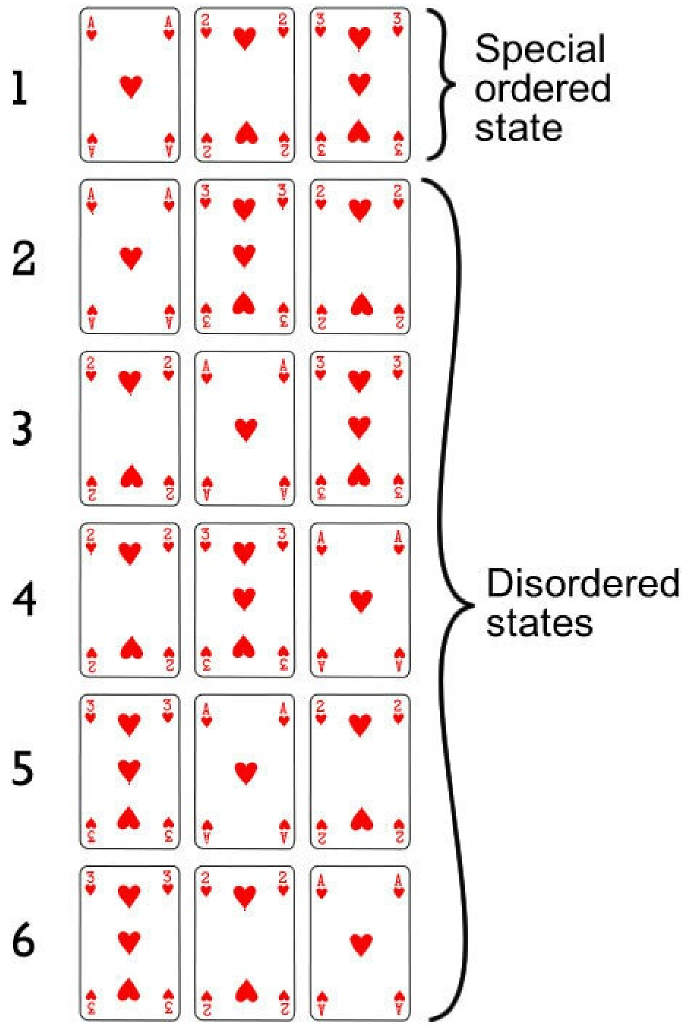
However, there is one example in which it might appear that life provides an exception to the inevitability of increasing entropy. Life has the ability to give birth to new life. And that new life, that new baby, appears perfect and untainted. So how can this apparent circumvention of the second law of thermodynamics be explained?

I do not believe the solution to this question which is generally presented is the correct solution. The usual explanation is that the second law only applies to closed systems, and a human being is not a closed system. A human interacts with its environment: it eats food, it expels waste. The food has grown under the Sun — itself a source of low entropy energy — and the expelled waste has high entropy. So the total entropy of a human system can be reduced through interaction with the environment.

However, as I say, I do not believe this is actually the correct explanation of how entropy is apparently massively reduced during the birth of a new human baby. In order to discover the correct solution, we need to consider *degrees of freedom*.

The number of degrees of freedom of a system is how many independent distinguishing values it possesses which are allowed to vary. That statement is not as complicated as it might sound. Once again, consider a pack of 52 playing cards which has been randomly shuffled. How do we identify one particular ordering of that pack of cards from the random ordering of a different pack of cards? Well, firstly, the two packs might have a different card as the top card in the pack. If the packs have a different top card, then that would certainly allow us to distinguish between the orderings of the two packs. However, if the two top cards are the same for both packs, we would then consider the second card in the packs. Basically, as I am sure you can see, there will be 52 different positions in the pack which will allow us to distinguish between the orderings of the two packs. We could therefore say that each pack has 52 degrees of freedom.

Now let us consider a pack consisting of only three cards: the ace, two, and three of hearts. This pack, therefore, has only 3 degrees of freedom, compared with the 52 degrees of freedom the pack had before. How many different random orderings of this three card pack are now possible? You can see from the following diagram that there are now only six possible different orderings:



One of the orderings — ordering number one at the top of the diagram — is the special, ordered state. But there are now only five other possible disordered states. This is a far fewer number of disordered states than was the case with the 52 card pack. This means the maximum *W* term (the number of possible disordered states) in Boltzmann's formula for entropy is smaller, and so maximum entropy is reduced.

Now let us return to consider the birth of a human baby. How is it possible that two, wrinkly, middle-aged, entropy-rich parents can create such perfect low-entropy offspring? Well, the degrees of freedom available for describing a baby are far fewer than those available for describing an adult human. Put simply, babies all tend to look quite similar to each other. They have fewer distinguishing features. It is not even possible to determine the sex of a baby from facial appearances.

The available degrees of freedom for describing the few cells of an embryo are even fewer. And — as with our three card pack of cards — fewer degrees of freedom means maximum entropy is reduced. Basically, it becomes impossible to produce a baby with high-entropy distinguishing features, such as wrinkles, or a moustache (unless you are very unlucky). Your baby is going to look pretty much like everyone else's baby. This explains how a ageing couple can produce perfect, low-entropy offspring.

The low entropy of the early universe is often presented as one of the great mysteries in physics. However, given our extraordinary knowledge of the state of the universe just a fraction of a second after the Big Bang, this surely should not be considered a mystery anymore. A simple comparison to a newborn low-entropy human baby can cast a light on this mystery.

Let us consider the universe and evaluate how many degrees of freedom it has. In how many ways could we distinguish between one state of the universe and a different state of the universe? Clearly, the number of different values needed to describe a universe is going to be an astronomically huge number. Hence, we could say that the universe has a staggeringly huge number of degrees of freedom.

But what about the first few seconds after the Big Bang? How many degrees of freedom would the universe have had then? It is likely that the universe would have had far fewer degrees of freedom. The universe was much smoother, being composed of pure energy, with the elementary particles yet to emerge. This embryonic universe resembled a smooth blob, marred only by tiny quantum fluctuations, with few distinguishing features: a true baby universe.

We normally associate a smooth distribution with high entropy, but this is not the case if the smoothness is an inevitable consequence of few degrees of freedom — as is the case with a baby. All babies are smooth, but all babies have low entropy.

So this reduction in the degrees of freedom provides a reason as to why entropy was so low in the early universe. As Roger Penrose explains in his book *The Road to Reality*: "There is a common view that the entropy increase in the second law is somehow just a necessary consequence of the expansion of the universe … There are comparatively few degrees of freedom available to the universe when it is 'small', providing some kind of low ceiling to possible entropy values, and more available degrees of freedom when the universe gets larger, giving a higher 'ceiling', thereby allowing higher entropies." (Penrose then continued to expound his own theory involving gravity, which I don't agree with).

As the elementary particles emerged, the maximum possible entropy massively increased. The particles became free to interact via **all** the fundamental forces. One millisecond after the birth of the universe, the strong nuclear force acted to clump quarks together to form protons and neutrons. Gravity acted to clump all particles to form stars and galaxies. Distinguishing features emerged within the universe. But the underlying reason for the low entropy at the very start of the universe was surely the very few available degrees of freedom.

Just as with the universe, a newborn baby is relatively smooth and featureless. However, as the baby matures, it develops its distinguishing characteristics. The baby develops more degrees of freedom. It develops clearer facial definition, it grows tall or short, it develops wrinkles.

All these characteristics are what help to distinguish us from each other. Society views wrinkles negatively, but it is experience with gives us depth and makes us interesting. And so it is with the universe. The universe was born smooth and perfect, but it only became interesting when the wrinkles (stars, planets, life) started appearing. We become more interesting as we age for precisely the same reason that the universe became more interesting.

Unfortunately, the second law suggests the entropy of the universe will continue to increase until it reaches a maximum value of disorder. At that point, all the atoms of the universe will be in a random state, so the universe will all be a constant temperature. At this point, the universe has run down. Nothing will happen in the universe. This fate of the universe has been given the name *heat death*.

To quote Jim Al-Khalili from *Order And Disorder*: "The process of change and degradation is unavoidable. The Second Law says the universe itself must one day reach a point of maximum entropy, maximum disorder. The universe itself must one day die."

Our journey through life matches the journey of the universe. We inhabit a universe which was born as a perfect baby, has its most interesting years during the time it ages and becomes wrinkly, before dying of old age.

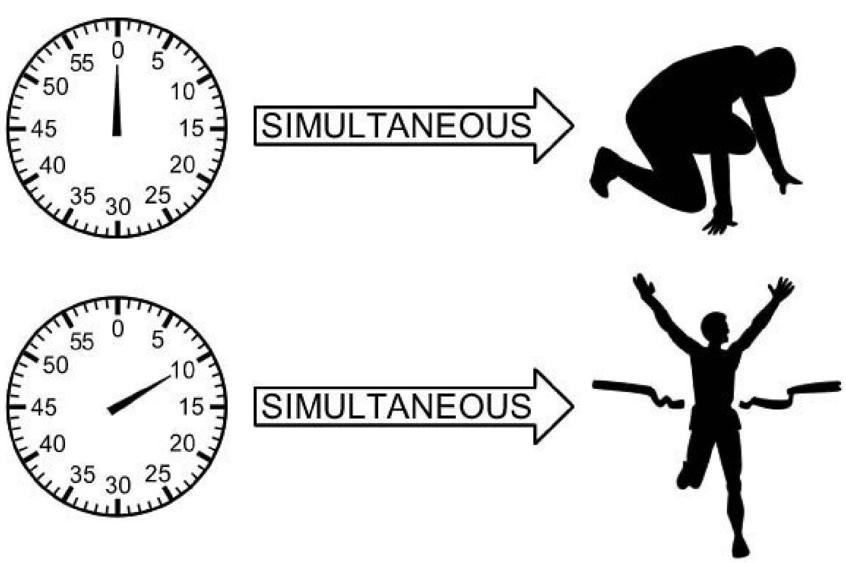
In this chapter we have seen it is impossible to define boundaries between ourselves and the external universe. We are all part of the one universe. So, in that sense, the universe is truly a living entity.

6

TIME AND RELATIVITY

If you remember back to Chapter Three you will recall a lengthy explanation of simultaneity: when events happen at the same time. It was explained how vital simultaneity was in shaping reality for different observers. In this chapter, I want to return to the subject of simultaneity and show how it is completely central to the topic of time.

If we think deeply about time and simultaneity, we can realise that all measurements of time are measurements of simultaneous events. For example, imagine I time a sprinter over a 100m race, and at the end of the race my stopwatch tells me 10 seconds have elapsed. What has actually happened in terms of time? Actually, all that has happened is that two pairs of simultaneous events have occurred. The first pair of simultaneous events was my stopwatch pointing to zero, and the race starting. The second pair of simultaneous events was the winning runner crossing the line and my stopwatch pointing to 10 seconds.



From the existence of these simultaneous events, we deduce something like: "The race took ten seconds", or "ten seconds of time passed". So we make it sound as if time is moving, or passing by, but really all we have is the existence of these two pairs of simultaneous events. There is nothing moving or dynamic here — there is just the existence of events.

Of course, you might want to sub-divide the race into smaller time intervals ("ticks") such as hundredths of seconds. But the same argument applies — you just end up with more simultaneous events. It is very much like frames of a movie: each frame of the movie represents a series of simultaneous events. If you have a sufficiently high frame rate then the scene appears to move. But all that really exists is a sequence of simultaneous events in each frame. All that really exists is the movie reel.

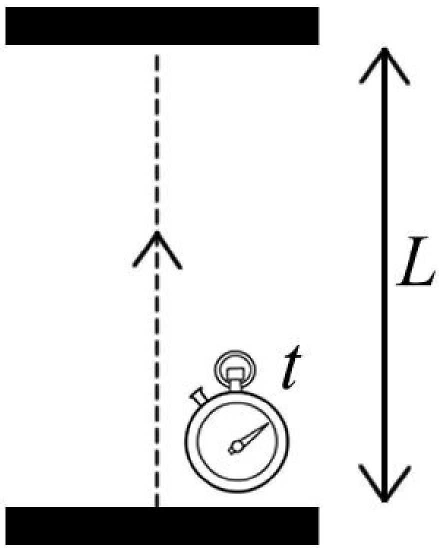
Einstein was well aware of this fact as he explained in his groundbreaking 1905 special relativity paper: "We have to take into account that all our judgements in which time plays a part are always judgements of simultaneous events. If, for instance, I say: 'That train arrives here at 7 o'clock', I mean something like this: 'The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous events.'"

So time itself is defined by simultaneous events. But, crucially, this means that anything that can affect the simultaneity of events can therefore affect time itself. Remembering back to our discussion of simultaneity involving Einstein's thought experiment about the moving train, it is clear just what can affect the simultaneity of events: motion! It was the motion of the train relative to the observer standing on the platform which affected which events Bob and Alice considered to be simultaneous.

So if simultaneity is affected by relative motion, then time itself must be affected by relative motion (remember: time itself is defined by simultaneous events). In that case, how might it be possible to generate quantitative equations to reveal precisely how much the passing of time is modified for moving observers? To calculate this, we will now follow Einstein's workings by using only high-school mathematics (this adds weight to my contention that, as we get to the most fundamental levels of Nature, the mathematics should become simpler — not more complex).

In order to generate his equation, Einstein considered the light clock which we first encountered in Chapter One. This consists of a ray of light bouncing between two mirrors which are a distance L apart (see the following diagram). This light clock was placed on a train travelling at a constant velocity. The clock was oriented so that the direction the light travels was perpendicular to the direction of motion of the train.

Bob is once again travelling on the train, and his job is to measure the time taken for the ray of light to travel between the two mirrors. This time — according to Bob — was measured as t:

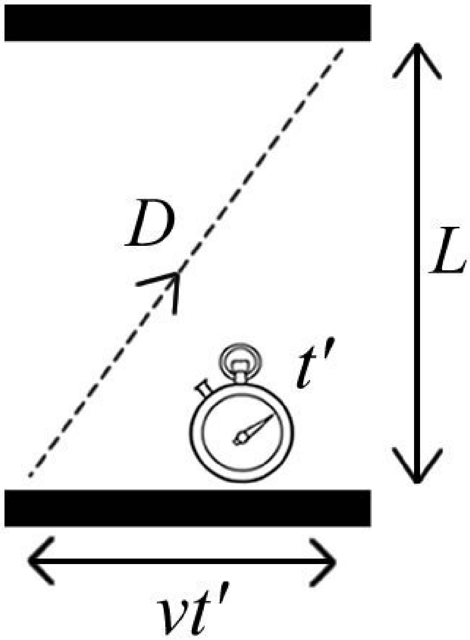


Then from distance = speed × time we get:

$$L = ct$$

However, to Alice — who sees the train moving — it appears the ray of light has to travel a greater distance. This is due to the additional velocity of the train making the light travel in a longer, angled path (see the following diagram). Alice on the platform measures this new time for the light to travel between the two mirrors to be t'.

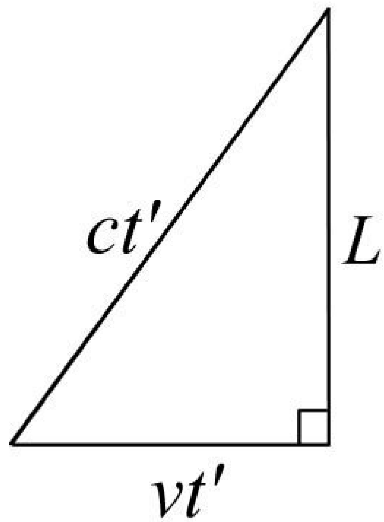
If the speed of the train is v, then the train travels a distance vt' in this time:



Referring to this second diagram, we see that the distance the light now has to travel (according to Alice) is *D*. Now, remember that all observers will measure the same value for the speed of light. So, according to Alice, the distance *D* is (from distance = speed × time):

$$D = ct'$$

If we consider the path of the light as forming the hypotenuse of a right-angled triangle:



We can then apply Pythagoras's theorem to this triangle (you will remember that Pythagoras's theorem says that the square on the side of the hypotenuse is equal to the sum of the squares on the other two sides). So:

$$(ct')^2 = L^2 + (vt')^2$$

From the earlier equation, we know that:

$$L = ct$$

So substituting this value for *L* into the previous equation gives:

$$(ct')^2 = (ct)^2 + (vt')^2$$

Which gives:

$$c^2 t'^2 = c^2 t^2 + v^2 t'^2$$

Rearranging to get all the t and t' values on different sides:

$$c^2t^2 = t'^2(c^2 - v^2)$$

Dividing all terms by c^2 gives:

$$t^2 = t'^2\left(1 - \frac{v^2}{c^2}\right)$$

Finally, taking the square root of both sides gives:

$$t = t'\sqrt{1 - \frac{v^2}{c^2}}$$

So here we have our final equation describing the phenomenon of time dilation. It shows that the rate that time passes for Bob on the train, t, is less than the rate that time passes for Alice on the platform, t'.

We can see from the equation that the rate at which time slows down for Bob on the train is given by:

$$\sqrt{1 - \frac{v^2}{c^2}}$$

So, if the train is travelling at 80% of the speed of light (it's an unusually fast train), we find time on the train slows down by a rate of:

$$\sqrt{1-0.8^2} = \sqrt{0.36} = 0.6$$

Which means that time for Bob on the train will pass at only 60% of the rate that time passes Alice on the platform.

Alice will actually age faster than Bob!

Interestingly, the formula for time dilation seems to indicate that the speed of light represents an ultimate speed limit for all objects in the universe (in the formula, if *v* was greater than *c* then we would have to take the square root of a negative number — and you can't do that). But why should the speed of light represent the ultimate speed limit? We will be returning to this question later.

The twin paradox

This principle of time dilation does seem to raise something of a puzzle. Consider a thought experiment involving two twins. One twin stays on Earth while the other twin gets in a spaceship and travels away at a speed close to the speed of light. When the spaceship turns around and returns to Earth, the twin who was in the spaceship finds that he has aged less than the twin who has remained on Earth. This is precisely in accordance with the expected time dilation.

However, the puzzle arises because surely all motion is relative. If that is the case it is just as valid for the twin on the spaceship to consider himself stationary, and the twin back on Earth to be moving at close to the speed of light. This principle of relative motion was expressed by Einstein who, when travelling by train from London to Oxford, asked the ticket inspector: "Does Oxford stop at this train?" Einstein was considering the train to be stationary, and the town to be moving.

So, if the twin on the spaceship considers himself to be stationary, surely he would expect the twin back on Earth to be the one who ages less? After all, it is now the twin on Earth who is doing the travelling.

This apparent paradox is called the *twin paradox*.

The resolution of the paradox comes from realising that the experiences of the two twins are not identical. To be precise, the twin on the spaceship experiences acceleration when he takes off, when he turns around, and when he returns to Earth. He would experience inertial forces due to this acceleration (we will be considering inertial forces in the next chapter). However, the twin who remains on Earth experiences no such acceleration.

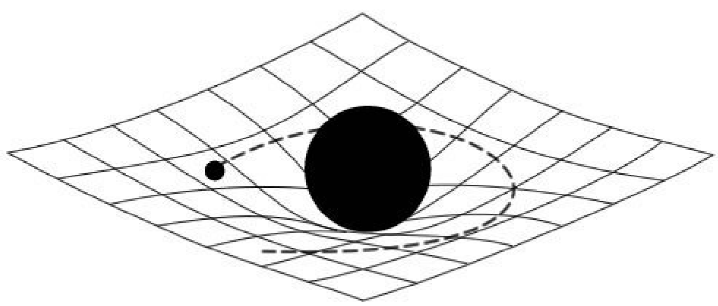
Once it is realised that the situation is not symmetrical, the formula for time dilation produces the correct non symmetrical result for the ageing of the twins. [7]

Gravitational time dilation

In this section we will consider a different type of time dilation predicted by relativity: the slowing of time in a strong gravitational field.

In 1915, Einstein published his revolutionary theory of general relativity. General relativity explained the force of gravity in terms of the curvature of space (or, more precisely, the curvature of spacetime). An object attempting to travel in a straight line in curved space would follow a curved path, and that deflection is interpreted as being due to gravity.

In a famous example, a large mass is placed in the middle of a rubber sheet, deforming the sheet in the same way that mass is known to curve spacetime:



A smaller mass will then appear to orbit the large mass in a circular path, although the smaller mass is really just trying to follow a straight line across a surface which is curved. As Einstein explained: "When a blind beetle crawls over the surface of a curved branch, it doesn't notice that the track it has covered is indeed curved. I was lucky enough to notice what the beetle didn't notice."

Because general relativity predicts the curving of space — the fundamental underlying structure of the universe — it predicts everything will be affected by gravity: even light. In the most extreme example, light is unable to escape from the intense gravitational field of a black hole. This is not due to "light slowing down" due to gravity — after all, light always travels at the speed of light. No, the light is trapped due to the extreme curvature of space at the black hole. Essentially, light would have to travel an infinite distance (due to the curvature of space) in order to escape from a black hole.

So the effect of a strong gravitational field is to curve space so that light has to travel a longer distance between two points than normal (just as a curved road between two points is longer than a straight route). However, we know that all observers have to measure the same speed of light. Therefore, if an observer is measuring the same speed for a light ray — even though it is having to travel a longer distance — then there is only one conclusion we can make: the observer's measuring clock must be running slow (we would obviously usually expect light to take a longer time to travel a greater distance). Hence, a clock runs slower in strong gravity.

This slowing of time in a strong gravitational field is called *gravitational time dilation*.

Hence, gravitational time dilation predicts that time will run faster at higher altitude (further away from the Earth's gravitational pull). With that thought in mind, let us return to the experiment we considered in Chapter One: the experiment in which Joe Hafele and Richard Keating flew four atomic clocks around the world. Because of the cruising altitude of the Boeing 747, gravitational time dilation predicted that the clocks would run slightly faster than clocks which remained at the U.S. Naval Observatory. The results were found to precisely match the predictions of general relativity, and were of similar magnitude to the additional time dilation predicted by special relativity (due to the speed of the aircraft).

Another example of gravitational time dilation is experienced by GPS (Global Positioning System) satellites. GPS satellites orbit 12,500 miles above the Earth's surface, hence our clocks on the surface of the Earth run slightly slower. The clock on the GPS satellite is therefore set to run 45 microseconds per day slower than usual in order to remain synchronized with clocks on Earth. Because of the nanosecond accuracy required by GPS, if this effect was not taken into account, positional errors would accumulate at the rate of ten kilometres per day.

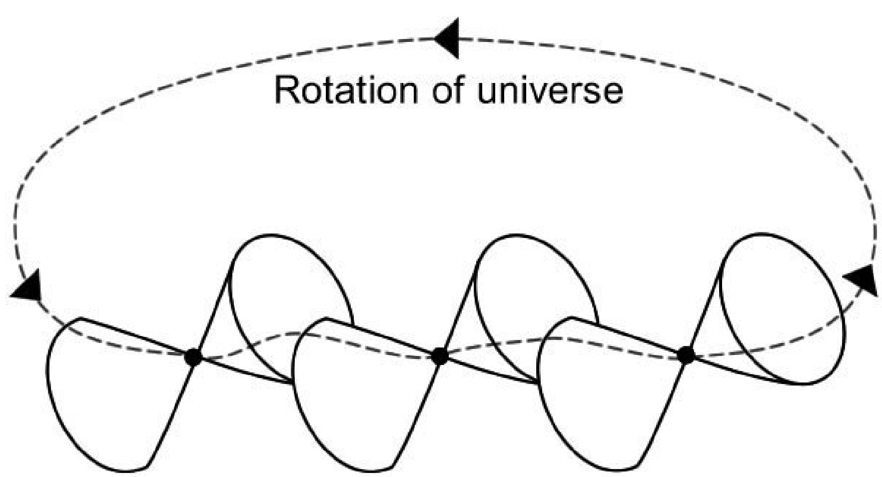
Time travel

The equations of general relativity describe how spacetime can be curved by the presence of mass. The equations are complex, but precise solutions to the equations have been found for idealised situations in which simplifying assumptions can be made. When Einstein was working in the Institute for Advanced Study in Princeton in the 1940s, one of his colleagues — and best friends — was the great mathematical logician Kurt Gödel (we will be meeting Gödel again in my next book, *Hidden In Plain Sight 4*). It was Gödel who discovered a solution to the equations of general relativity which first revealed the possibility of time travel to the past.

Gödel's solution was based on the phenomenon of *frame-dragging*. Frame-dragging occurs when a large mass rotates. In that case, general relativity predicts that spacetime in the immediate vicinity of the mass will be dragged around in a circle.

Gödel considered the frame-dragging which would occur in the hypothetical situation of the whole universe rotating. In that case, all of the spacetime in the universe would be dragged around, much like the rotational mixing of a sticky substance in a food mixing bowl. In that case, the future could be bent around so that it connected to the past. By travelling far enough into the future along your timeline, it would then be possible to travel to the past.

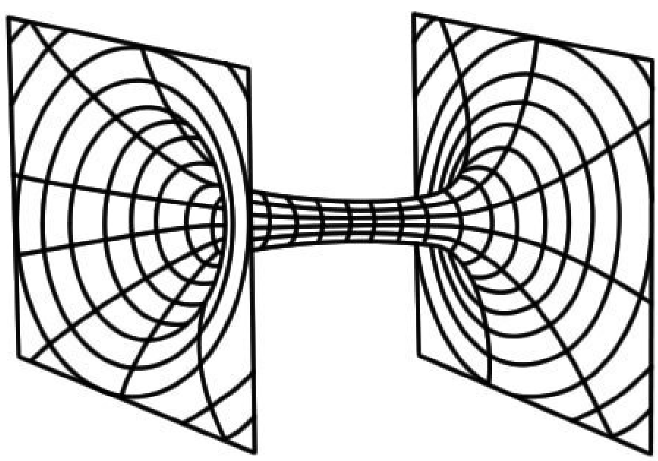
The following diagram (based on a diagram from Paul Nahin's book *Time Machines*) shows how future light cones would be tipped over by frame-dragging, so that the future can point into the past:



It is possible to use the frame-dragging phenomenon to design a time machine. In 1974, the American physicist Frank Tipler published quite specific design details for a time machine.[8] No one had ever published such a paper in a prestigious physics journal before. Tipler's time machine proposal required a long, very dense cylinder rotating with a surface speed of at least half the speed of light. The resultant frame-dragging around the cylinder would result in a similar time loop to the one generated in Gödel's universe. As Tipler says at the end of his paper: "In short, general relativity suggests that if we construct a sufficiently large rotating cylinder, we create a time machine."

No one disputes Tipler's claim — it really is a blueprint for a time machine.

Another time machine proposal comes from a solution to the general relativity equations by Albert Einstein and Nathan Rosen. Their solution resembled a tube effectively providing a shortcut between two regions of spacetime. This was called the *Einstein-Rosen bridge*, though John Wheeler popularised the name *wormhole*:



By creating a shortcut in spacetime, wormholes raise the possibility of travelling great distances in space in a very short time. This possibility was explored in Carl Sagan's novel *Contact* (and later film of the same name). In the book, a radio telescope in New Mexico receives a coded transmission from extraterrestrial beings. The decoded message forms instructions to build a vehicle which is used to transport five passengers through a wormhole to a planet near the distant star Vega — in merely a few seconds. After spending several hours on the alien planet, the travellers return to Earth to discover that only a few seconds have elapsed on Earth since their voyage began.

This plot hints that wormholes can also be used for travelling in time as well as space. Indeed, assuming that it is possible to create a traversable wormhole, it would be a fairly straightforward procedure to turn that wormhole into a time machine. The following procedure was described by Kip Thorne in his book *Black Holes and Time Warps*. A wormhole has two mouths and there is no obvious connection in space between the two mouths: if I step through one mouth of the wormhole, I immediately emerge out of the other mouth. With no obvious connection between the two mouths, the two mouths can be moved quite independently of one another. To turn our wormhole into a time machine, one mouth must be kept on Earth, while the other mouth is sent away on a spaceship at a speed approaching the speed of light (this is treating the wormhole mouth very much like one of the twins in the previous thought experiment). Time dilation ensures that time passes more slowly for the wormhole mouth travelling on the spaceship. So when the spaceship returns to Earth, if you walked through the wormhole on the spaceship, you would return to an earlier time.

All these time machine proposals raise problems as they introduce the possibility of changing the past. If it is possible to change the past then that could potentially lead to the *grandfather paradox*. The grandfather paradox considers the possibility that you use a time machine to travel back in time to kill your own grandfather before you are conceived. However, if as a result of you murdering your grandfather, you are not conceived then you would not have been able to travel back in time. Hence, this results in a paradoxical situation.

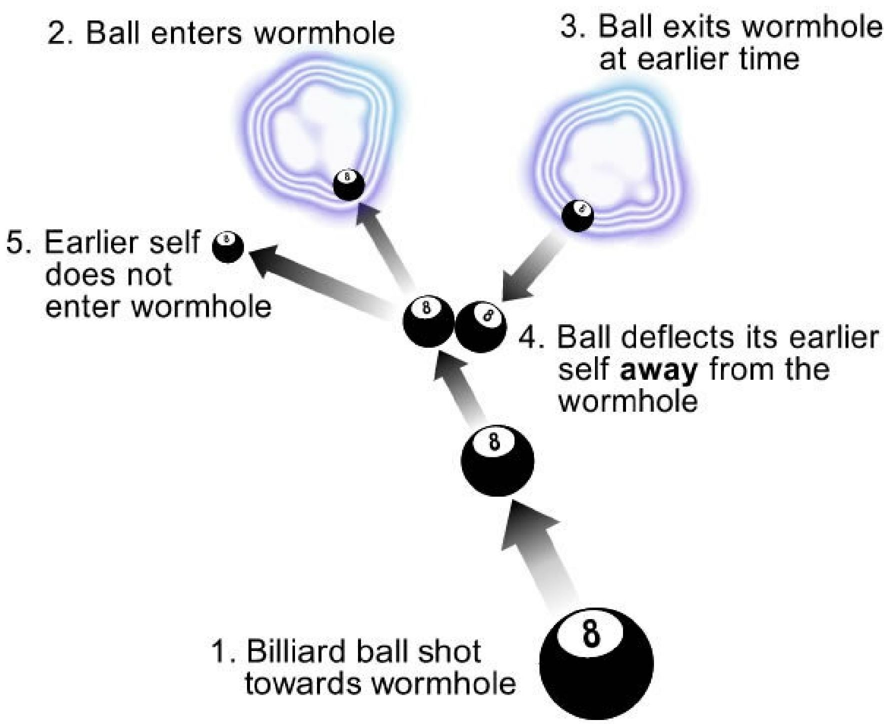
From the point of view of this book, the ability to change the past poses a particular problem. All through this book, the block universe structure has been presented as a static, unchanging block of spacetime. There is no possibility of altering any aspect of the structure — it is as if it is carved in stone. However, if it is possible to travel back in time to change the past then this appears to disprove the whole block universe argument.

Fortunately, as we shall now see, the solution to the grandfather paradox also solves the apparent problem with the block universe model.

Another form of the grandfather paradox was proposed in 1990 by Joe Polchinksi, then a professor of physics at the University of Texas. What was ideal about Polchinski's paradox was that it avoided any questions about human free will (would you really be able to kill your own grandfather?).

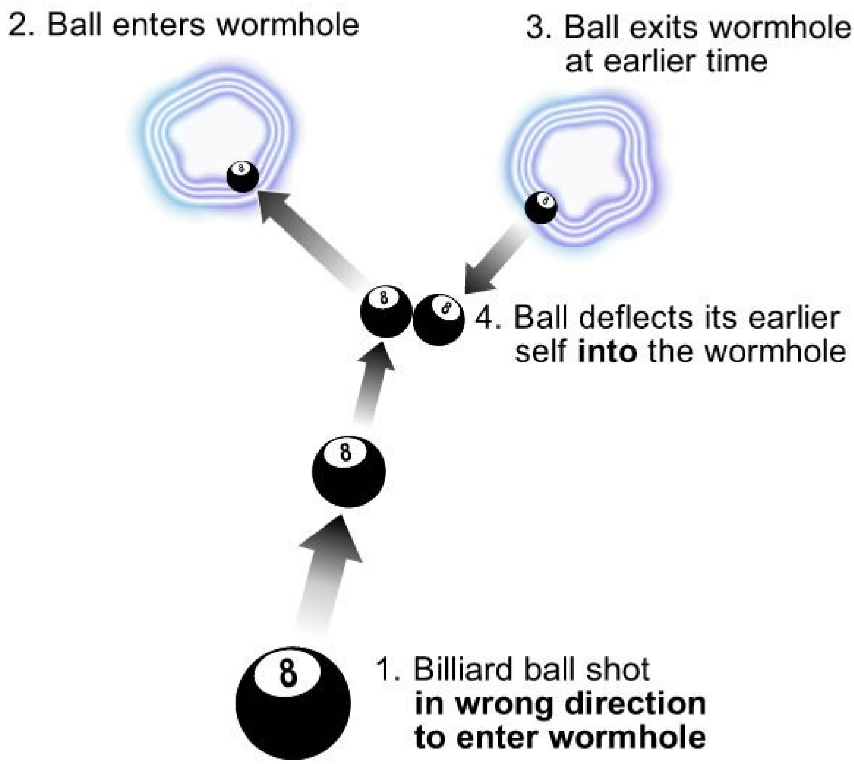
Hence, the paradox could now be analysed purely from a physics viewpoint.

The reason Polchinski's paradox did not involve human free will was because it did not involve humans. Instead, it involved billiard balls. The paradox considered a billiard ball rolling along a billiard table and entering a wormhole mouth at the far end of the table. It is then possible for that billiard ball to travel back in time and emerge out of the second wormhole mouth. Polchinski's paradox considered the situation in which the second wormhole mouth was also positioned on the billiard table. What would happen if, when the billiard ball travels back in time and emerges out of the second wormhole mouth, it collides with its earlier self and deflects its earlier self away from the wormhole? It would appear the ball is prevented from entering the wormhole in the first place. This would result in a paradox very similar to the grandfather paradox. If the billiard ball is prevented from ever entering the wormhole because it is deflected, then it cannot emerge from the second wormhole mouth and deflect itself.



It would appear that this scenario would raise a clear paradox. It would also not be reconcilable with the block universe model. It would appear that such a scenario could not occur in reality. Does this mean wormholes cannot exist? Or that the block universe model is wrong?

Fortunately, it has been realised that it is possible to travel back to the past — and even affect past events — without introducing paradoxes. Consider the previous billiard ball example again. When the ball travels back to the past and emerges from the wormhole, it is possible that it deflects its earlier self **into** the wormhole, instead of **away** from it:



There is now no paradox in this result. It is also a situation which can exist in a block universe: the past is affected, but events which have already occurred (i.e., set in stone in the block universe) have not be altered.

This principle in time travel — that only situations can occur which do not result in paradoxes — is called the *Novikov self-consistency principle*.

7

INERTIA

In this chapter we will be considering motion. There is an inextricable link between time and motion. Motion is defined as the variation in position of an object at different times. So without time, there can be no motion. It could also be argued that without motion, there can be no time. As Richard Wolfson describes in his book *Simply Einstein: Relativity Demystified*: "The study of motion is profound, for several reasons. First, motion is the source of all change. Imagine a world without motion: Earth stops rotating, so it's perpetual daytime. Atoms cease moving, so there's no chemistry — no release of energy. Nothing evolves, transforms, mutates, develops, or otherwise changes. What does it mean to move? It means getting from one place to another, and doing so in some time. Whatever else motion means, it involves passing through time and through space. **So motion holds the key to understanding time and space.**"

In this chapter we will be paying attention to a particular kind of motion called *inertial motion*. Inertial motion is the motion of an object which is not being acted-on by any force. The resultant inertial motion is actually very mysterious. For example, if you were travelling at constant velocity on a very smooth train, you could consider yourself to be in inertial motion. As a result, as long as the train is smooth enough, you would not feel as if you were travelling at all! So this clearly raises some fascinating and important questions about the nature of motion, and what it means to be "moving".

Inertial motion seems to be telling us something very profound about the nature of time and space. In the next chapter, it will become clear as to why inertial motion is so important for our understanding of time.

Perpetual motion

The Greek philosopher, Aristotle, wondered if there was a "natural state" of motion. As Richard Wolfson again explains, the natural state of motion would be a state which requires "no explanation — a state that an object naturally assumes unless something is explicitly done to it, like pushing or pulling it." To this end, Aristotle considered moving objects, such as a block of wood being pushed along a table, and saw that all such moving objects eventually came to a halt. Hence, Aristotle believed that any moving object would eventually slow down and come to a halt — unless it was acted upon by a constant force to keep it in motion.

From this observation, Aristotle believed he could deduce the natural state of motion. As Aristotle observed that it required a force to maintain an object in motion, he believed the natural state of motion was to be stationary. A stationary object is to be expected — it requires no explanation. Whereas if you see a moving object, you could reasonably seek an explanation, an answer to the question "What is making it move?"

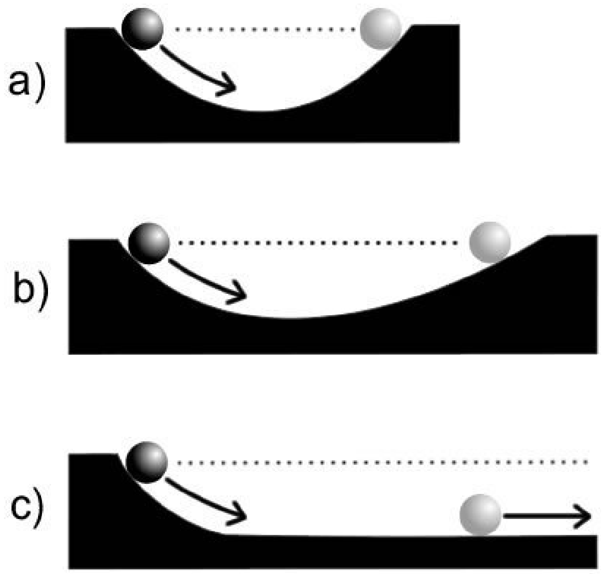
Aristotle's argument regarding the natural state of motion was convincing, and remained dominant until Galileo performed his experiments of motion. According to Galileo: "Ignorato moto, ignorator natura", a Latin phrase which translates to "He who fails to understand motion, fails to understand Nature." We considered the principle of Galilean relativity in Chapter Two. It is Galilean relativity which introduced the world to the principle of inertial motion.

If you remember, Galilean relativity states that the laws of motion are the same for all observers who are moving at a constant velocity. There is no experiment you could perform to determine if you were stationary or moving at constant velocity.

It is this movement at constant velocity which is inertial motion. If it is impossible to distinguish between being stationary and being in inertial motion, then clearly an observer moving with inertial motion will feel stationary (otherwise it would be possible to detect the motion). So Galilean relativity reveals that an observer moving at constant velocity (inertial motion) will feel stationary.

It is the principle of Galilean relativity which is the reason why the geocentric model of the universe proved dominant for so long (rather than the heliocentric model being adopted). This is because the early astronomers were fooled by the undetectable inertial motion of the Earth. Because the Earth orbits the Sun at a constant, smooth velocity, and because the astronomers moved at the same rate as the Earth, they were fundamentally unable to detect its motion: they felt stationary. Indeed, the notion that the Earth orbits the sun would have appeared an absurd suggestion to them. [9]

Galileo then considered Aristotle's concept of a natural state of motion. If you remember, Aristotle believed that the natural state of motion was to be stationary, as a moving object would tend to become stationary. However, Galileo performed an ingenious series of experiments which led him to a different conclusion. These experiments of Galileo are illustrated by the following diagrams:



As shown in Figure a), Galileo rolled a ball down a slope and discovered it rose to the same height up an opposite slope. As shown in Figure b), if Galileo increased the distance between the slopes, the ball still rose to the same height up the opposite slope — but it obviously had to travel further. As shown in Figure c), Galileo then reasoned that if there was no opposite slope, the ball would travel forever as it would be forever unable to reach the initial height up the slope.

Galileo's conclusion was remarkable. It suggested that an object moving horizontally in a straight line would continue indefinitely. At this point, you might raise an objection. Surely motion cannot continue indefinitely without continually providing some energy to the system? In raising this objection, maybe your instincts have been biased by some of the crazy designs for perpetual motion machines which are all doomed to failure.

It is certainly true that it is not possible to build a practical perpetual motion machine. Such a device would crank round a few times before seizing to a halt. The reason why all these machines fail is because they do not take into account the force of friction. Friction converts kinetic energy into heat energy which is radiated from the system. This loss of energy results in the inevitable slowing of the system.

However, if it could be possible to eliminate the force of friction from the system then the motion of the machine would indeed continue forever — as predicted by Galileo. For example, flywheels have been built which float on superconducting magnets (virtually frictionless) and are sealed in a vacuum to eliminate air resistance. These flywheels will keep spinning for a period (the so-called zero-load rundown time) of many years. If such a device could ever be made completely friction-free then it would, indeed, be a perpetual motion machine. In fact, electric currents (the flow of electrons) in superconducting materials can continue in motion indefinitely — a form of perpetual motion which is already with us.

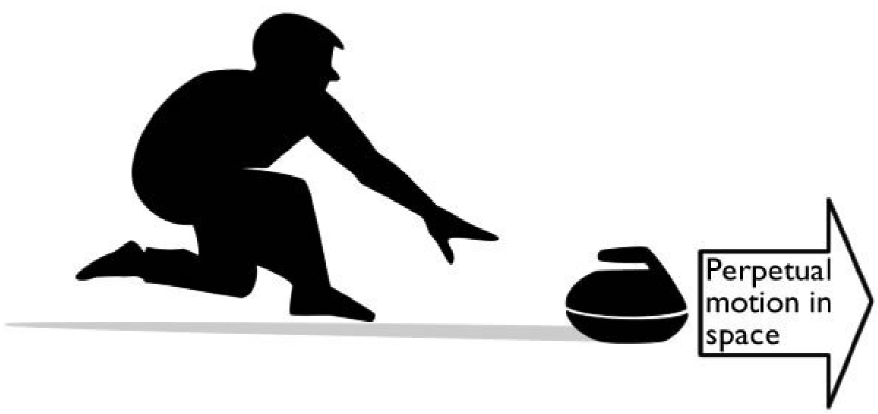
The notion of perpetual motion might seem very wrong and alien to us. Our worldview has been so shaped by all the tales of doomed perpetual motion machines that we instinctively discard any notion that perpetual motion — of any kind — might be possible. On the basis of this discussion, however, we find that **perpetual motion is not only possible, it is natural**.

Earlier in this section we considered Aristotle's argument that the natural state of motion was to be stationary — as it required no explanation. Now, however, we see that Galileo proved Aristotle to be wrong. As Richard Wolfson states in his book *Simply Einstein*, Galileo showed that "motion itself needs no cause or explanation … motion is natural." According to Galileo, the natural state of motion is no longer being stationary — the natural state is motion at a constant speed in a straight line, in other words the natural state of motion is inertial motion. This is a hugely important result, and we will return to it later in this book:

The natural state of motion is not being stationary, it is constant velocity in a straight line (inertial motion).

This principle is called *inertia*. The discovery of inertia was Galileo's greatest contribution to physics.

You have probably seen the Olympic sport of curling, in which a granite stone is sent sliding across ice. The stone travels at low speed, but the almost frictionless ice ensures the stone hardly slows down at all. This is an excellent example of inertia. If we could imagine a perfectly frictionless form of ice then it is clear that the stone would travel at its low speed forever:



Galileo died in 1642, the same year that Isaac Newton was born. When Newton started as a student in Trinity College in Cambridge University he found that the curriculum was still dominated by the thinkings of Aristotle. Newton introduced the college to the new developments of Copernicus and Galileo.

Newton accepted Galileo's conclusions about the laws of motion, and built on them himself to produce his three laws of motion which form the basis of classical mechanics:

1. Every object in a state of uniform motion tends to remain in that state of motion (or remain at rest) unless an external force is applied to it.

2. Forces act to accelerate objects. The acceleration is proportional to the force which is applied, and inversely proportional to the mass of the object: $F=m\times a$

3. For every action there is an equal and opposite reaction.

Newton's first law of motion is just a restatement of Galileo's concept of inertia, and so the first law is often called the "law of inertia". The word "inertia" comes from the Greek word for "laziness", and we can see where this comes from. The definition of inertia is that an object likes to continue in its state of motion — be that moving or remaining stationary — unless a force acts on it.

Newton's three laws of motion defined our conception of mechanics for the next three hundred years.

Inertial mass and gravitational mass

Einstein's great talent was for identifying a simple, unifying link between two seemingly distinct effects. This talent was clearly shown when he realised that the force due to gravity was indistinguishable from the force due to acceleration. This led to the theory of general relativity (for details, see my previous book). This realisation allowed Einstein to unify two properties which, for centuries, had been considered to be distinct: inertial mass and gravitational mass.

The *inertial mass* of an object determines its resistance to acceleration, i.e., its inertia. Newton's second law of motion says that the acceleration experienced by an object is inversely proportional to the inertial mass of that object.

It might seem strange to us now, but after Newton published his laws of motion and his law of gravitation, it was considered that the inertial mass of an object might well be different from the mass of the object which attracted other objects: the *gravitational mass*. After all, the two effects — acceleration and gravity — were considered to be completely distinct effects. There was no obvious reason why the object might not behave differently when subjected to the two different forces.

Of course, when Einstein realised that the force due to gravity was completely equivalent to the force experienced during acceleration this removed the centuries-long distinction between gravitational mass and inertial mass. An object is now considered to have just one single value for its mass, and this is used for calculations of both acceleration and gravity.

However, experiments have continued to test if any difference between gravitational and inertial mass can be detected. In 1885, the Hungarian physicist Roland Eötvös (pronounced "urt-vursch") proposed the most famous of these experiments, an experiment which is now known simply as the Eötvös experiment. The experiment uses a balance (called a *torsion balance*) on which masses are dangled on a piece of string which can rotate. The outward force on the masses (due to centrifugal force) would be dependent on the inertial mass. The downward force on the masses is dependent on the gravitational mass. The masses are carefully selected so that the string will not rotate if the inertial mass is precisely equal to the gravitational mass so that the effects cancel each other. The torsion balance at the University of Washington is so sensitive that it can tell if it has rained recently as the extra water in the soil has an increased gravitational pull on their equipment.

But there is a simpler way to test if inertial mass is equal to gravitational mass. Newton's laws of motion only deal with inertial mass, the second law stating that a force applied to an object will produce an acceleration according to $F=m_i \times a$, where m_i is the inertial mass. On the surface of the Earth, for a falling object, we can equate this force to the force of gravity:

$$m_i a = \frac{GMm_g}{R^2}$$

Where m_g is the gravitational mass, M is the mass of the Earth, and R is the radius of the Earth. If inertial mass equals gravitational mass then $m_i = m_g$, and the two terms cancel from the equation. We are left with a formula for the acceleration due to gravity on the surface of the Earth:

$$a = \frac{GM}{R^2}$$

There is no term in this formula which is dependent on the mass of the falling object — the only mass featured in this equation is the mass of the Earth. So this indicates that — if inertial mass really is equal to gravitational mass — all masses will fall at the same rate (just as Galileo demonstrated on the Leaning Tower of Pisa centuries ago).

So it is possible to test the equivalence of inertial and gravitational mass just be checking if all masses fall at the same rate, and this test can be performed with great accuracy. The Center of Applied Space Technology and Microgravity (ZARM) in the University of Bremen is dominated by a 146-metre-tall drop tower which provides 9.3 seconds of free-fall during which experiments can be performed. Many different fields of research use the weightless environment. Various masses have been dropped to see if they all fall at the same rate, and no discrepancy has been found:

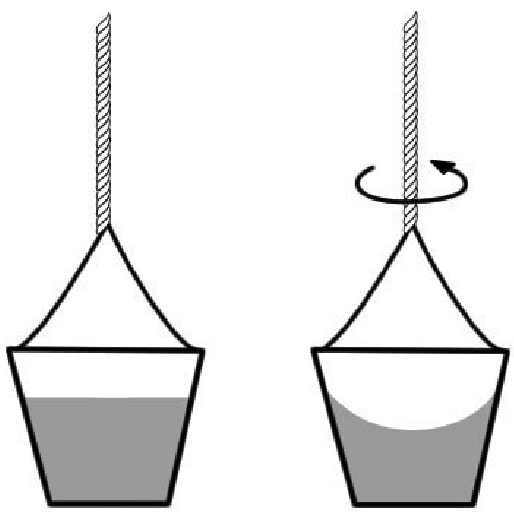
In all these experiments, the inertial mass and the gravitational mass of various elements has been found to be identical to 13 decimal places. Almost certainly, no difference will ever be found.

Inertial forces

Consider a question, the answer of which might appear quite obvious at first glance, but is not as simple as it might appear. The question is: how do you determine that an object is moving? If you look out of your window, for example, you might see a bird flying past. It would appear easy to tell that the bird is moving. But why should the bird not consider itself to be stationary, and it is you who are moving? In other words, if motion is relative, how is it possible to tell what is stationary and what is moving?

In this case of the bird, it would appear that it is possible to tell that the bird is moving relative to the trees and buildings around it. In other words, both the bird and the observer could agree that the bird is moving relative to its **immediate environment**. So motion relative to the immediate environment would appear to be a way to definitively determine if an object is moving.

However, Newton raised an objection to this apparent importance of the immediate environment. He described an experiment in which a bucket containing water was spun at speed at the end of a rope. The water starts to spin within the bucket:



If all motion was relative, then we would be justified in considering the water to be stationary and the rest of the universe revolving around the water. However, the water does not behave as if it is stationary — the water surface becomes concave and climbs up the sides of the bucket (see the second bucket in the previous diagram). Somehow Nature can tell that the water is moving.

From our previous discussion about the bird outside your window, we can see that the obvious way to determine that the water is moving is to consider its motion relative to its immediate environment. However, the immediate environment of the water is the sides of the bucket, and the water is not moving at all relative to the sides of the bucket (the water is rotating at the same speed as the bucket). So the principle of using the immediate environment to determine the relative motion of an object is not applicable in all cases.

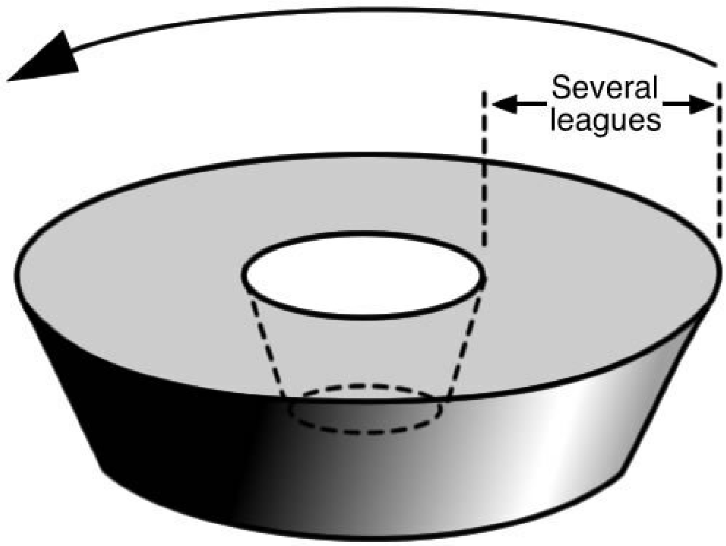
Newton took this result as being evidence for absolute space. According to Newton, absolute space permeates the universe and provides an absolute reference for motion. Nature can determine that the water is rotating relative to absolute space, and this is the reason the water rises up the sides of the bucket.

Newton's bucket argument was so convincing that absolute space was accepted as the true model of space for 200 years, until an alternative was provided by the Austrian physicist Ernst Mach in the late 19^{th} century.

In his career, Mach contributed much to physics. You will recall the speed of supersonic aircraft being referred to as "Mach 1", "Mach 2". Those terms are named after Ernst Mach whose experiments in supersonic velocity included the definition of the *Mach number* — the ratio of a projectile's velocity to the speed of sound. But it is Mach's insights into the nature of space — and the cause of inertial forces — which interests us here.

Mach considered the conclusions Newton drew from his bucket experiment and decided Newton was too quick to disregard the effect of the environment. True, the water was not rotating with respect to the immediate environment, but there were many more objects in the environment than just the bucket. For example, there is the planet Earth, or the planets, or the fixed stars in the sky. And these distant objects would possess vastly more mass than the sides of the bucket. Maybe they were more influential? Maybe Nature could detect that the water was rotating relative to this more distant, hugely more massive environment?

So Mach considered the situation in which the sides of the bucket were much larger, more distant, possessing hugely more mass:



What would happen now when this massive bucket rotates with the water? As Mach wrote in *The Science of Mechanics* (1883): "No one is competent to say how the experiment would turn out if the sides of the vessel increased in thickness and mass until they were ultimately several leagues thick."

It is now believed that, if the sides of Mach's bucket really were sufficiently huge, spacetime would be dragged around inside the bucket at the same angular velocity as the water. As a result, the water inside the bucket would, indeed, **not** rise up the walls of the bucket. This is the frame-dragging effect (a phenomenon described in the previous chapter on time machines). NASA's Gravity Probe B satellite was sent into orbit in 2004 to see if it could detect frame-dragging of spacetime by the rotation of the Earth. The experimental equipment consisted of four extremely sensitive gyroscopes — the most perfectly spherical manmade objects ever created. Essentially, the Earth performed the role of Mach's huge bucket, and the four gyroscopes on the satellite performed the role of the water in the bucket, detecting any evidence of frame-dragging. In 2011 it was announced that the experiment had successfully measured frame-dragging due to the rotation of the Earth — the first time the effect had been detected.

So Mach believed that motion of an object could be defined relative to all the other masses in the universe, and this was the cause of inertia — the reason you feel a force when you are accelerated. Nature can detect the change in your velocity relative to all the other objects in the universe. As Mach was quoted as saying: "When the subway jerks, it's the fixed stars that throw you down."

Einstein was heavily influenced by the work of Mach. As with Mach, Einstein was also not convinced by Newton's concept of absolute space. However, this idea of Mach's — of the fixed stars billions of miles away having an instantaneous effect on your velocity — sounded far too much like Newton's instantaneous action-at-a-distance theory of gravity.

So Einstein proposed a gravitational field which spread throughout space. This gravitational field controlled the motion of masses. An object in free-fall would follow the field lines of this gravitational field. If an object was accelerated, it would cross the field lines of the gravitational field (essentially the definition of acceleration — a deviation from free-fall motion). Nature could then determine an object was being accelerated.

This, then, provides an explanation for the inertial force you experience when you are accelerated. The physicist John Stachel has coined the term *inertio-gravitational field* for this reason. Just as Einstein revealed the equivalence between gravity and acceleration, and the equivalence between inertial mass and gravitational mass, so the term "inertio-gravitational field" stresses that just the one field is responsible for both inertial and gravitational forces.

John Woodward, in a paper entitled *What is the Cause of Inertia?*, is in no doubt: "What is the cause of inertia? Gravity." [10]

Imagine you are a Grand Prix motor racing driver driving at speed around a tight corner. Your car — and your body in the cockpit — changes direction around the corner, and therefore it is deviating from the normal free-fall direction of travelling in a straight line. This means the car is crossing inertio-gravitational field lines: it is accelerating. However, your head is sticking out of the cockpit and wants to continue in free-fall inertial motion in a straight line down the road. The stress this causes in your body — as if your head is being pulled off — is the inertial force felt during acceleration.

The force the driver feels is often described in terms of *g-force*, maybe 3g, 4g, or 5g. A force of 1g is defined as being equal to the force due to the force of gravity, so the concept of g-force again reveals the link between gravity and inertia.

So the sideways force you feel as you drive around a corner at speed is actually the force of horizontal gravity!

But what determines the shape of the inertio-gravitational field? The curvature of the gravitational field is inevitably determined by all the other masses in the universe. After all, the distant stars may be billions of light-years away, but we are still affected by their gravity — as weak as it may be. All the masses in the universe are inevitably affected by the gravitational pull of all the other masses in the universe. So the strength and direction of the gravitational field at any point is determined by the distribution of masses in the rest of the universe.

In this way, general relativity is very much Machian in nature. In the absence of absolute space, the inertial forces you feel are inevitably controlled by the position and distribution of the fixed stars because these celestial bodies determine the shape of the inertio-gravitational field throughout the universe. Of course, the greatest influence on the gravitational field in our immediate vicinity is provided by the planet Earth. And the resultant force of gravity we feel directed towards the centre of this planet is usually the greatest inertial force we experience.

Unless you are a Grand Prix driver.

8

THE UNIVERSAL SPEED

In the course of the previous chapters, we have gained a working understanding of the phenomenon of time dilation. In Chapter Six we derived the formula to determine how time slows down for a moving object. This was based on the principle of relativity, the principle that all the laws of physics should be the same for all observers. In turn, that implies that the measured speed of light should be the same for all observers. We have noted that this reminds us of the Copernican principle: no observer holds a privileged position in the universe. No observer is special — all observers are equal. This seems very much like a fundamental principle which would necessarily be true.

However, there are several indications that we have not yet got to the bottom of the mystery of relativity. Specifically, the centrality of the speed of light in all this seems rather puzzling. Why should electromagnetic radiation (light) play a role in determining the passage of time? Why does the speed of light appear to represent an absolute speed limit for all objects in the universe? Is there perhaps some deeper principle at play here which we need to uncover?

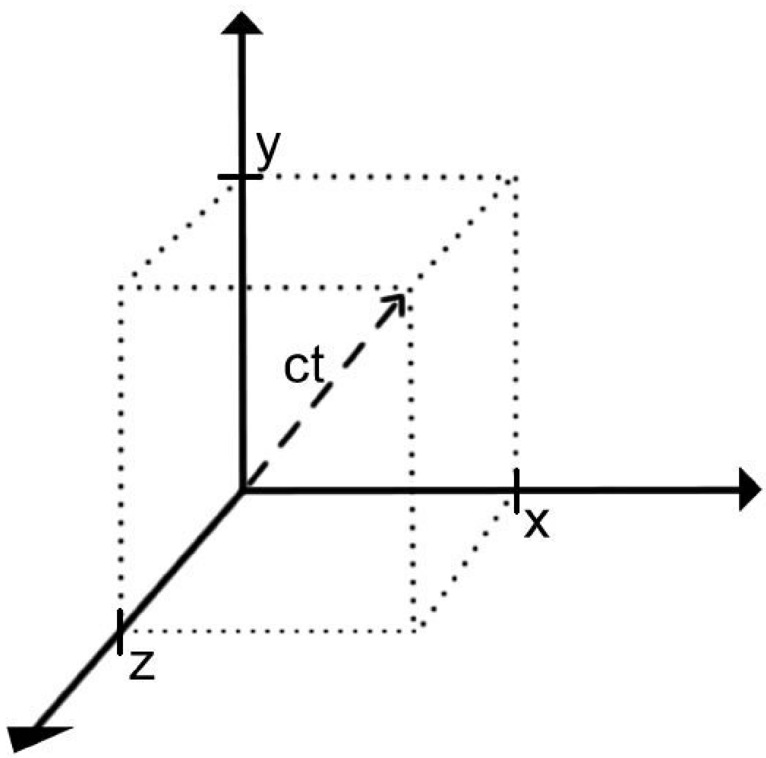
In this chapter, we start to get some answers …

The first thing we need to consider is how we might go about creating a mathematical model of spacetime. If we manage to create a mathematical model then we can analyse it and hopefully gain a deeper understanding of its properties.

But how on earth are we to combine measures of space and measures of time in order to create a model of a combined "spacetime"? To quote Brian Cox and Jeff Forshaw from their book *Why Does E=mc*^{2}: "If distance in space is measured in meters and distance in time in seconds, how can we even begin to contemplate combining the two. It is like adding apples and oranges, because they are not the same type of quantity."

The one piece of absolutely vital information we can use is that all observers — no matter how they are moving — must measure the same value for the speed of light. So, as speed represents a connection between distance and time, this could potentially provide the means for combining space and time into spacetime.

Imagine there is a flash of light at the origin of our coordinate system. In the following diagram, the path of light is shown by the dashed arrow coming from the origin.



The sphere of light in three-dimensional space will expand outwards and, after a period of time *t*, will have a radius of *ct*. Considering the diagram, and using Pythagoras's theorem extended to three dimensions, we get:

$$x^2 + y^2 + z^2 = (ct)^2$$

which gives:

$$x^2 + y^2 + z^2 - (ct)^2 = 0$$

The expression on the left hand side of this equation is very important. Because all observers will agree on the measured speed of light, all observers will agree on the value of this expression. Also, we see the expression combines measurements of both space and time. So this is a combination of space and time on which all observers can agree. The square root of the value of this expression is given the name *spacetime distance* (or the *spacetime interval*), though this "distance" value actually combines both space and time.

Now, this is where things start to get really interesting. Because, as described in the previous chapter, every observer in inertial motion considers himself to be perfectly stationary — it is all the other objects which are moving relative to you. Essentially, in the previous diagram, this means you are permanently placing yourself at the origin of the coordinate system and everything else in the universe is defined relative to your position. In the previous formula for spacetime distance, this means that your values for x, y, and z are all zero (because you do not move relative to yourself). But, if you consider the formula for spacetime distance, you will see that if x, y, and z are all set to zero then the value of the distance in spacetime does not become zero — it becomes equal to ct.

What does this mean? It means that even if you are perfectly stationary, the spacetime distance you travel is equal to ct, which is the speed of light multiplied by time. This can only mean one thing: even when you feel stationary, **you are moving in spacetime at the speed of light!** In fact, in the next section we shall see that everything moves through spacetime at the speed of light.

At last, this is a remarkable result which seems to reveal a deep truth about the nature of time. As Brian Cox and Jeff Forshaw say in their book *Why Does $E=mc^2$*: "The statement that everything moves at the same speed through spacetime sounds rather profound."

Euclidean relativity

It might seem surprising to hear that all objects move through spacetime at the speed of light. After all, we obviously see various objects (cars, birds, etc.) moving around us at different speeds. However, crucially, note that these objects are moving at different speeds in **space** — not in **spacetime**. Once time is taken into account, and we consider the combined speed of these objects in a spacetime composed of both space **and** time, then we find that everything is moving at the same speed. As Brian Greene explains in his book *The Fabric of the Cosmos*: "The combined speed of any object's motion through space and its motion through time is always precisely equal to the speed of light."

How can this be? How can considering time as well as space make all the difference? Well, it is because of time dilation: an object which is moving relative to an observer will experience less time. So, to put it simply, **an object which is moving faster through space will move slower in time**. If an object is moving at a fixed speed — the speed of light — in spacetime then this effect is inevitable. After all, there is only enough speed to go round. If an object uses up the majority of its speed to travel through space, then it has less speed left over to travel through time.

As Brian Cox and Jeff Forshaw explain: "This newfound way of thinking about how things move through spacetime can help us get a different handle on why moving clocks run slow. In this spacetime way of thinking, a moving clock uses up some of its fixed quota of spacetime speed because of its motion through space and that leaves less for its motion through time."

So the principle that everything travels at the same speed in spacetime predicts time dilation: a reduction in measured time for a moving object. And if we calculate the amount of time dilation predicted by this principle we find it is exactly the same amount predicted by special relativity. However, this new, simple principle is surely the true principle which lies behind relativity.

This innovative approach to relativity is called *Euclidean relativity* (although Carl Brannen calls it *proper time geometry* in a paper on the topic [11]). I am surprised it does not get more recognition.

But, of course, this is not the whole story. The truth is surely that the true fundamental principle behind relativity is the Copernican principle. In spacetime, no point is preferred, no observer is special, and all observers move at the same speed in spacetime. Every object is treated exactly the same. Every object moves at exactly the same speed.

In human society, rich people can move faster by buying expensive cars and planes. However, physics recognises no such inequality. In physics, everyone — and every object — is equal. Everyone is treated the same. The laws of physics are the same for everyone. Everyone ages the same as entropy takes it toll equally. Under the laws of physics, everyone moves at the same speed in spacetime. You cannot buy a faster car to drive through spacetime more quickly. This is the true essence of the egalitarian Copernican principle.

The world of physics is a more equal world.

Sliding through time

At the start of this chapter, a couple of relativity-related questions were posed. It was wondered why electromagnetic radiation (light) should play a role in determining the passage of time. Also, it was wondered why the speed of light should appear to represent an absolute speed limit for all objects in the universe. It was suggested that there was maybe a more fundamental principle that we were missing that might provide the answers.

Well, now we have our more fundamental principle, and now we have our answers. **Everything** moves at the same speed in spacetime — not just light. There is nothing special about light. Light does not play a central role in determining the passage of time.

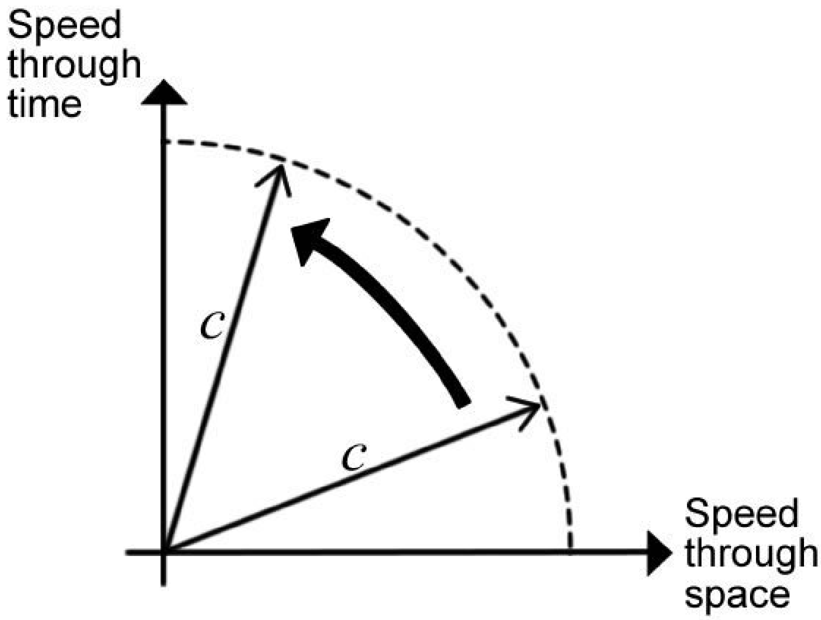
So why does light feature in most descriptions of special relativity? This is because light is unusual in that it has no mass. This means that light is free to travel at the maximum speed through space (mass normally acts to restrict speed through space). Hence, all of light's speed through spacetime is directed through space, and none of its speed is directed through time. As a result, light does not experience the passing of time.

This means that light's speed through spacetime is very obvious to our eyes — it is the same as its speed through space. So out of all the objects in the universe, only light appears to be travelling at this magical universal speed. Whereas, in fact, everything is travelling at this universal speed — in spacetime.

So the speed of light is not a universal speed limit at all, it is just the speed that everything is moving — in spacetime.

With this insight, I would suggest we no longer refer to objects travelling at the speed of light but instead we should refer to the *universal speed* — the speed at which everything travels in spacetime.

Consider the diagram below which shows an object slowing down while travelling in space. Initially, its speed arrow in spacetime is pointing to the right, showing that most of its speed is through space. However, as the object slows, its speed arrow in spacetime rotates anticlockwise so that less of its speed is through space, and most of its speed is through time:

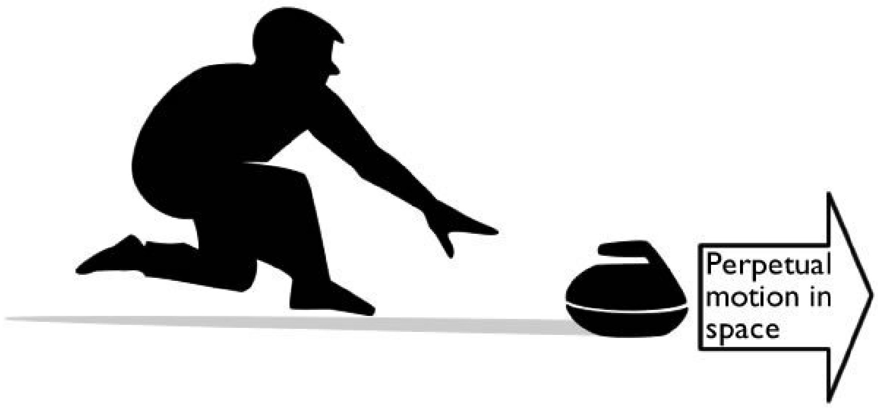


So as you see objects slowing down or speeding up, what is actually happening is that their speed arrow in spacetime is rotating. But the actual speed of the object through spacetime is unaffected (the length of the speed arrow remains unchanged).

Now, this is interesting. Perhaps we can obtain further insights from this approach.

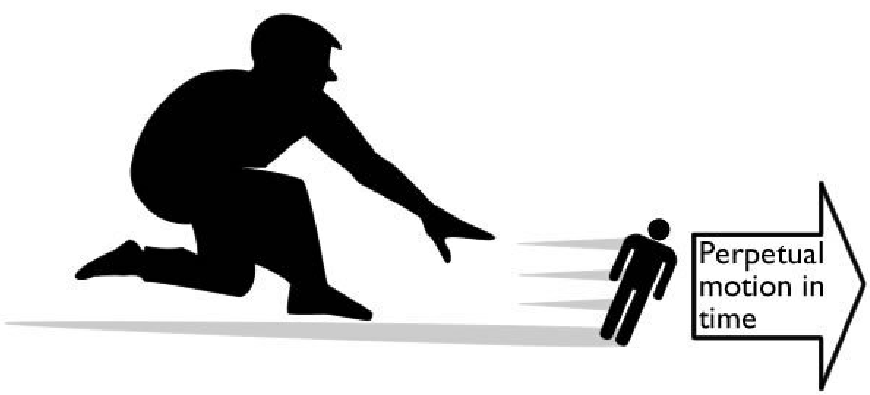
The principle that all objects move at the same, constant speed in spacetime is interesting. If they are moving at the same constant speed — never slowing down — then this sounds very much like a form of perpetual motion. Even when objects appear to be slowing down, all that is really happening is that their speed arrow is rotating in spacetime. And where have we heard about perpetual motion before? Well, in the last chapter on inertia it was revealed that the natural state of motion is perpetual motion. We now see this is true: the natural state of motion is perpetual motion in spacetime.

In the previous chapter, the example was presented of curling: if the ice was completely frictionless then the stone would continue its perpetual motion forever.



However, as the stone eventually slows (due to friction) its speed arrow rotates in the anticlockwise direction. Eventually, the stone appears to stop. However, the magnitude of the speed arrow in spacetime is unaffected, and the stone therefore continues to move **in time** at the universal speed. So what has really happened is that the inertial, perpetual motion of the stone in space has rotated to become inertial, perpetual motion in time.

So this gives us a tremendous insight. The motion we feel in time — the passage of time — is caused by our inertial motion in time! The principle which is causing our motion in time is the same principle which ensures the curling stone slides forever across frictionless ice. It is as if we are sliding through time!



Just as the perpetual motion of the curling stone across the frictionless ice is natural motion which requires no additional source of power, so our perpetual motion through time requires no additional power. It is the natural state of motion.

9

ENERGY AND MOMENTUM

Throughout this book, it has been found that by considering the motion of objects we have obtained insights into the nature of time. From considering curling stones on ice, to trains on railway tracks, we have discovered the similarity — and symmetry — between motion in space and motion in time. In this chapter we are going to consider two more related properties of objects in motion: their energy and momentum. We will see if these properties can provide further insights into the nature of time.

It was Rene Descartes and Isaac Newton who first identified the property of momentum of a moving object. Momentum is the property you get when you multiply the mass of an object with its velocity. For example, a heavy lorry moving fast would have a lot of momentum. So the formula for momentum is mass multiplied by velocity: mv.

Momentum does not just have a numerical value, it always has an associated direction. Essentially, this means that momentum is always represented by an arrow (an arrow has both a magnitude (length) and a direction). The correct technical term for such an arrow is a *vector*. Momentum is a vector quantity.

Perhaps the most interesting feature of momentum is that it is always conserved, which means the total momentum of a closed system does not change with time. In collisions, momentum can be transferred from one object to another, but if you add up all the momentum of all the objects you will find the total momentum before the collision is equal to the total momentum after the collision.

This principle that momentum is always conserved leads us to think that momentum represents some important underlying feature of Nature. We will see later just how important this is.

Momentum is even conserved in situations in which all the objects are initially stationary — such as an explosion. Imagine a stick of dynamite standing in a stationary position on a table. Nothing is moving, so the total momentum of that system is zero. When the dynamite explodes, fragments will shoot out at great speed to the left and right, and up and down. So it might appear that momentum is not conserved here because a stationary situation has transformed into a situation with a great deal of motion. However, remember that momentum is a vector quantity: it is represented by arrows. The momentum of a fragment which shoots out to the left can be represented by an vector of a certain length pointing to the left. Similarly, the momentum of a fragment which shoots out to the right can be represented by a vector of a certain length pointing to the right. If we add a vector pointing to the left with a vector pointing to the right then they cancel each other out and the sum total is zero. So if we correctly consider all the momentum of all the exploding fragments as vectors shooting out in all directions, we will find the sum total is again zero. So total momentum is conserved even in the case of an explosion.

(This principle — that momentum must be conserved in an explosion — is a big clue. Because the Big Bang can be considered to have been an explosion, so momentum must have been conserved over the period of the Big Bang. What does that imply? We will return to consider this idea later.)

But, here is an interesting thought. Imagine a car driving along a road at a constant speed. It clearly has momentum. It also clearly has an amount of kinetic energy, the energy associated with movement. We could tie a rope to the back of the car and, as the car pulls on the rope, it could make a windmill spin. So we could use the energy of the car to do work. This all makes sense and probably fits with your intuitive notion of momentum and energy.

However, as we discussed in the second chapter, we also know that all motion has to be relative — there is no such thing as absolute motion in the universe. With this in mind, imagine you are driving alongside the first car, and you choose a constant speed which precisely matches the speed of the first car. Now, as you look across to the first car, it no longer appears to be moving: it appears stationary. In fact, as far as you are concerned, the first car **is** stationary. With no observer having precedence in the universe, your viewpoint is just as valid as the viewpoint of any other observer. So, as far as you are concerned, the first car now has no momentum, and it has no kinetic energy. For example, if you have a windmill on your car, you could no longer use the energy of the first car to turn the blades of your windmill — the attached rope would be slack.

So this reveals that the momentum and kinetic energy of an object are completely observer-dependent, which might come as something of a surprise (it is something they don't tell you in school!). This also seems to point to a possible connection between momentum and energy, which we will discover shortly.

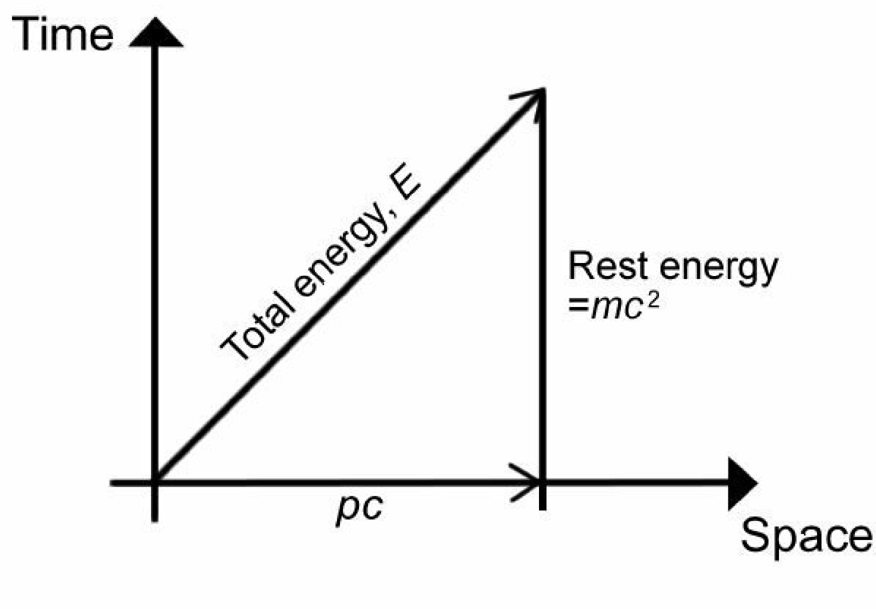
However, as we discussed in the previous chapter, everything travels at a constant speed in spacetime: the speed of light. Motion in space is really just a shadow of true motion in spacetime. So this observer-dependent impression of the momentum and energy of the car in space is not the whole story. To see the whole story we have to consider the motion of the car in spacetime — not just in space.

Once again, imagine you are driving alongside the first car, and you are driving at precisely the same constant speed as the first car. As you look across, the first car appears stationary. In fact, in your frame of reference, the first car **is** stationary. But, as we discussed in the previous chapter, we know that everything moves at the speed of light in spacetime. So what has effectively happened here is that — as far as you are concerned — the spacetime velocity vector of the car has rotated so that it no longer points in space, but now points only in time. In other words, the first car — which you see as stationary — must now be moving at the speed of light through time. This provides us with a general result: **a stationary object moves through time at the speed of light**. The only reason the object appears stationary to our eyes is because we are travelling through time at exactly the same speed!

But momentum is associated with motion (remember: momentum is mass multiplied by velocity). Admittedly, momentum is usually only associated with motion through space. However, in our new spacetime model, we might now reasonably ask what is the momentum of the car through time. The answer, fairly obviously, is that momentum is the mass of the object multiplied by its speed through time. So for a stationary object, momentum through time is given by the mass of the object multiplied by the speed of light: mc.

Now, this is interesting. The momentum through time of a stationary object is mc. What other property of a stationary object do we know? Well, Einstein showed in his famous equation $E=mc^2$ that the total energy of a stationary object (its *rest energy*) is equal to mc^2. This indicates that if we consider the momentum of an object through time, and multiply that value by the speed of light, we obtain the total energy of that object. This seems to indicate that **we can consider energy to be momentum in the time direction**.

We can show this relationship between momentum and energy in a spacetime diagram. The following diagram takes into account an object's motion in both space and time. It has just been explained how we can consider energy to be momentum in the time direction, so the rest energy of a stationary object (mc^2) is drawn in the vertical direction. Momentum in space is drawn in the horizontal direction:



The total energy of a moving object is a combination of that object's motion through space (its momentum) and its motion through time (its rest energy). On the diagram, you can see that this total energy, *E*, is a combination of those two types of motion, and is shown by the diagonal arrow. It is clear that we can calculate the length of the arrow from Pythagoras:

$$E^2 = (mc^2)^2 + (pc)^2$$

This is a hugely-important equation called the *energy-momentum relation*. It reveals the relationship between energy, mass, and momentum, and it applies to absolutely everything in the universe, from a car driving along a road to the energy and momentum of light itself.

Let us consider some examples of the use of this equation. Firstly, if we consider a stationary object which has mass — for example, a block of wood on a table — that object has no momentum in space so $p=0$ in this case, and the equation reduces to $E=mc^2$, the famous formula for the energy of a stationary object. So that is correct.

Let us now consider the other extreme, an object which is moving at the maximum speed through space: the speed of light. So what moves at the speed of light? The answer is in the question: light moves at the speed of light. Let us consider the particles which make up light which are photons. Photons are massless, which is the reason light can travel vast distances from the stars. So, for a single photon, $m=0$ in the equation, and the equation then reduces to $E=pc$, which is known to be the correct relationship between energy and momentum for a photon (yes, even though photons are massless they still have energy and momentum).

So the energy-momentum relation is correct for the two extremes, and for everything else in between.

Why does $E=mc^2$?

Let us now consider what we have discovered about the nature of energy, and, in particular, what we have discovered about $E=mc^2$.

$E=mc^2$ is perhaps the most famous equation in physics. As the amount of mass in this formula is multiplied by the incredibly huge value of the square of the speed of light, this formula reveals that a small amount of mass can be converted to a tremendous amount of energy. This is the principle behind nuclear power and nuclear weapons. As Andrew Steane says in his book *The Wonderful World of Relativity*: "This means that the total daily energy production of all the power stations in the world could in principle be obtained from just 14 kilograms of raw material."

How are we to make sense of this formula indicating that there is an equivalence between mass and energy? Some puzzling questions might include "How can an amount of mass at rest contain such a huge amount of energy?", and "Why does the speed of light feature so prominently in an equation linking mass and energy?"

Well, armed with the insights we have gained so far, we can now provide answers to these questions.

In the previous chapter it was shown that all objects travel at the speed of light in spacetime. This principle implies that all the spacetime speed of a stationary object must be directed through time: a stationary object moves through time at the speed of light.

So this, then, provides the answer to the mystery of $E=mc^2$. A stationary object is, in fact, far from stationary: it is speeding through time at the speed of light. It is this tremendous momentum in the time direction that results in a mass at apparent rest containing a huge amount of energy. As Richard Feynman said, this is "an energy that a particle possesses from its mere existence." [12]

And this reveals the secret as to why the speed of light appears in the famous equation linking mass with energy.

So, by considering all objects as moving at the speed of light through spacetime, we are shedding light on numerous mysteries. As Carl Brannen says in his aforementioned paper: "Various odd attributes of the theory of relativity, such as the huge amount of energy present in matter, and the impossibility of matter exceeding the speed of light, become natural consequences of a universal speed for all matter and energy."

10

THE SECRET OF TIME

The aim of this series of books is to try to uncover some fundamental answers, to get to the bottom of things, to find out "why" things are the way they are. The aim is to discover how much of the universe could have been created differently ("contingent"), and how much is a logical necessity. Our eventual goal is expressed by Einstein: "What I am really interested in is whether God could have made the world in a different way; that is, whether the necessity of logical simplicity leaves any freedom at all."

Regarding the laws of Nature, if a law was found to be a logical necessity for the existence of the universe, then it would appear that there could be no choice of the form that law could take. In other words, if the universe could only exist with a law of Nature taking a particular form — with no possible alternative allowing the existence of the universe — then it would appear that that would necessarily define the form of the law.

With this thought in mind, I believe it is possible to show that the universe could not exist without a dimension of time. Hence, time would be shown to be a logical necessity.

In order to arrive at this original hypothesis, we follow the method used in all my books: building-up from fundamental principles. These fundamental principles have to be so axiomatic, so obviously correct, that they would have to be true in any conceivable universe. Such a principle has to be self-contained, containing within itself the reason why it is obviously true.

It appears it is possible to build-up a universe from a remarkably small number of such principles. My list of fundamental principles is small, but it is growing. These principles are:

- The second law of thermodynamics. The principle that disorder will tend to increase. As we have seen in this book, this is the principle behind the arrow of time.

- The principle that "there is nothing outside the universe". If the universe is the totality of everything that exists (which it is — by definition), then this principle is obviously true. If you read my first book then you know that this is a remarkably powerful principle, potentially explaining symmetries in Nature, and quantum mechanical and relativistic behaviour. It also leads to …

- The principle that the universe has zero total energy. This principle played a central role in my second book, which showed it potentially led to a intriguing modification of general relativity.
- The Copernican principle. The principle that no point in the universe is preferred, no observer is special. It has been suggested in this book that this is the reason why all objects travel at the same speed: the speed of light in spacetime.

In this book, we have just found a remarkable result: momentum in the time dimension is responsible for energy. This is quite a shocking twist. Who would have suspected that two apparently completely unrelated quantities, time and energy, would be so closely related? This is suspicious. We seem to have stumbled upon evidence linking time to a completely different property. Is this the clue we have been seeking?

Let us see if energy is referenced in our list of fundamental principles. Yes, the third principle in the list says that the universe must have zero total energy. You will be aware of this principle if you have read my second book. As Misner, Thorne, and Wheeler said in their classic textbook *Gravitation*: "There is no such thing as the energy (or angular momentum, or charge) of a closed universe, according to general relativity, and this is for a simple reason. To weigh something one needs a platform on which to stand to do the weighing."

There can be no such weighing platform outside the universe — because there is nothing outside the universe. You could never put the universe on weighing scales to determine the total energy of the universe.

If you read my second book you will be aware that it is possible to attain a universe with zero energy if gravitational energy is considered negative. How can gravitational energy be negative? Well, if objects are separated to infinity they feel no gravitational pull between themselves, so the gravitational energy of the system is zero in that case. But when those objects were initially clumped together, you had to put a lot of energy into the system to force them apart. So if you have to put energy into a system just to get to a zero energy situation, this means the energy of the system when those objects were initially clumped together must have been negative.

This energy is called the *gravitational binding energy*, and is equal to the amount of energy which would be required to separate a group of masses to infinity. In a page on his excellent website, Matt Strassler calls this energy *interaction energy*, and makes the crucial point that interaction energy can be negative. [13]

So now we see why we need time: **time gives us energy!** Our emphasis on time in our investigation has been a red herring: the secret was nothing to do with time, **it was all about energy all along!** What a twist! Time was only involved because momentum in the time direction gives us energy. A time dimension is required so that energy is possible.

A zero-energy universe satisfies the vital principle that energy is conserved over the period of the Big Bang. As Alan Guth explains in *The Inflationary Universe*: "If the creation of the universe is to be described by physical laws that embody the conservation of energy, then the universe must have the same energy as whatever it was created from. If the universe was created from nothing, then the total energy must be zero." Basically, what this means is that the universe must have had zero energy before the Big Bang (as absolutely nothing was in existence), so the universe must also have zero total energy in the era after the Big Bang.

A universe made of energy is a universe which can sum to zero. A universe made of energy is a universe which can exist.

The antimatter connection

It is clear that time plays an important role in the existence of the universe. Let us now further examine the important role played by time by considering a few very simple thought experiments. These will reveal remarkable and simple logical connections between time, electric charge, energy, and the remarkable substance known as antimatter. Even though it might appear that the concepts of time, charge, energy, and antimatter are very complicated, the extraordinary simplicity of the following thought experiments will show that the connection is actually surprisingly simple and logical.

Firstly, let us consider two electrically-charged particles: a proton (which is positively charged) and an electron (which is negatively charged). The rule is that opposite charges attract, so in the following diagram we see motion of the electron towards the proton:

Now let us imagine exactly the same situation but with the direction of time being reversed (for example, we might have recorded a video of the previous example and are now playing the video in reverse):

As you can see in the previous diagram, with time being reversed, we see the motion in the opposite direction: we will now observe the electron moving away from the proton. Where there used to be an attraction, there is now a repulsion.

This might appear to be trivial and obvious because it is so simple, but actually it reveals a very fundamental symmetry between time and electric charge: it reveals that if we reverse the direction of time — and also reverse the electric charge of a particle — then the situation will be unchanged. [14]

So if we reverse the direction of time, we can apparently change a negatively-charged electron into a positively-charged electron. And that is very interesting, because we know that a positively-charged electron already exists and has been detected in experiments: it is called a *positron*. A positron is an example of *antimatter*. It is believed that every particle has a corresponding antiparticle which has opposite charge. The antiparticle is then an example of antimatter.

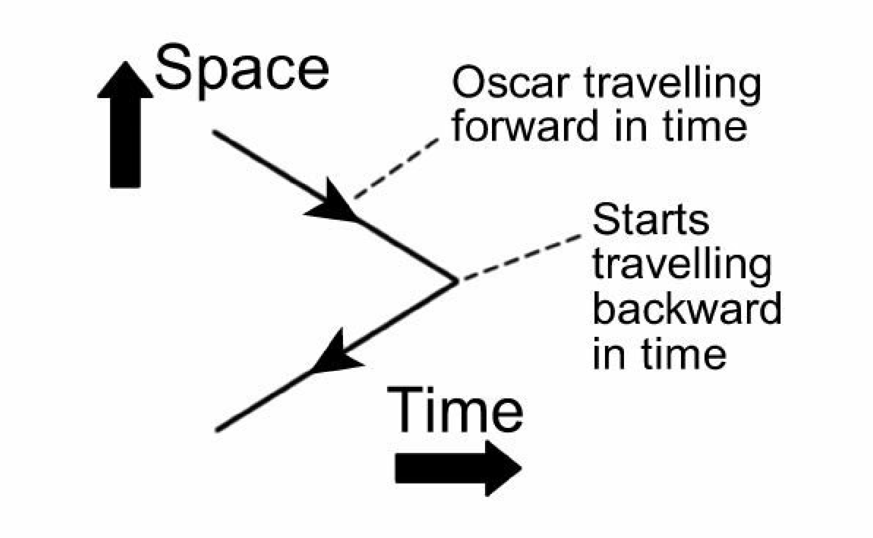
This principle (of reversed time producing antimatter) is described by Lisa Randall in her book *Warped Passages*: "Imagine a movie of a current of negatively-charged electrons travelling from one point to another. Now imagine running the movie in reverse. Negative charge would then travel backwards, or equivalently (so far as charge is concerned), positive charge would travel forwards. A current of positrons, the positively-charged antiparticles of electrons, produces this positively-charged forward-travelling current and therefore acts like a time-reversed electron current."

Antimatter is used as a common theme in science fiction, but it is most certainly a real substance and is even produced naturally in radioactive decay (a banana will produce one positron every 75 minutes or so). However, antimatter is very rare — and for good reason: when antimatter comes in to contact with normal matter they destroy each other, releasing a tremendous amount of energy in the form of gamma rays (high-energy light). In fact, antimatter releases more explosive energy than anything else in existence (an antimatter bomb would be 1,000 times more powerful than the atomic bombs dropped on Hiroshima and Nagasaki).

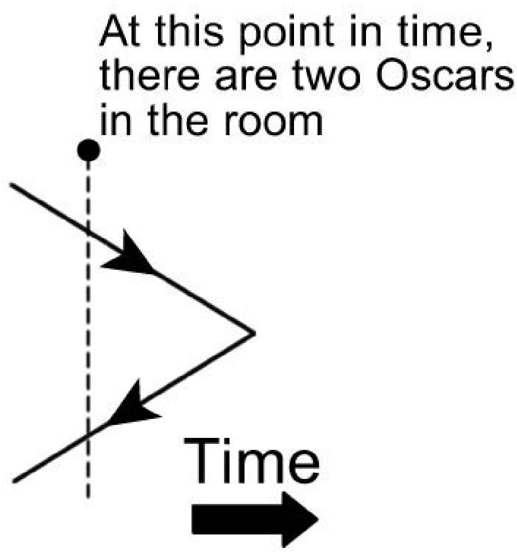
Now let us perform another simple thought experiment, which will reveal another highly-surprising connection between antimatter and reversed time. In this experiment we will discover just why matter and antimatter annihilate each other.

The following example is taken from a 1962 paper in the *Journal of Philosophy* written by the American philosopher Hilary Putnam.[15] The paper considers the strange adventures of a time traveller called Oscar Smith.

Let us examine the situation in more detail. The following diagram is a spacetime diagram showing the path of Oscar in space and time. We can see Oscar travelling forward in time, as he walks across the room. At a certain point, he starts travelling backward in time (the hands on his watch start to turn in reverse, for example) as he continues to walk across the room:



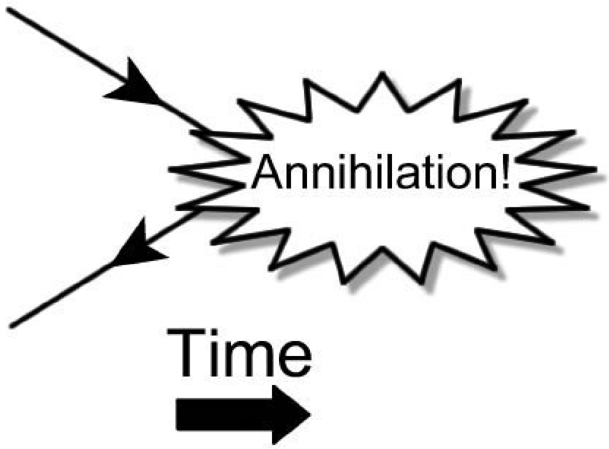
If we are an external observer, what would we see? Well, clearly — and remarkably — there would be a point in time (shown by the dotted line in the following diagram) when we observe two Oscars in the room! One of the Oscars would be travelling forward in time, and in a different position in the room we would observe the second Oscar travelling backward in time:



It is crucial to note that although there are two Oscars in the room at that point in time, they are in fact both the same man!

Now, we know from our previous analysis that a particle which is travelling backward in time is the antimatter equivalent of that same particle travelling forward in time. So can we consider the second Oscar — travelling backward in time — to be the antimatter equivalent of the first Oscar who is travelling in the forward time direction? Let us continue our analysis to see if this is the case.

What else do we observe in the room? Well, as time progresses, we observe the two Oscars moving closer together (the second Oscar moving backward in time, walking in reverse toward the first Oscar). You will then see from the spacetime diagram that when the two Oscars eventually meet, they do not appear to exist on the spacetime diagram after that point in time. It is as though they both vanish into thin air!



As Hilary Putnam says in his paper: "When the two systems merge, they are both annihilated".

So this does, indeed, resemble the behaviour of antimatter. Remember that when antimatter comes into contact with normal matter they both annihilate each other and cease to exist.

So, rather wonderfully, purely by considering two simple thought experiments we have managed to discover a connection between antimatter and reversed time, and we have found a simple explanation for the apparent annihilation of antimatter and normal matter. And I can assure you that this all represents orthodox scientific thinking (treating antimatter as the time-reversed equivalent of normal matter was first suggested by the great physicists John Wheeler and Richard Feynman).

It might be imagined that the connection between an exotic material such as antimatter and the direction of time, and the reason why antimatter annihilates normal matter would be an incredibly complex question. However, as we have seen, we can deduce the connection from some incredibly simple reasoning. I think this lends weight to my belief that — as we get to the lowest, fundamental levels of physics — we find physics gets simpler, not more complicated.

These insights into the behaviour of antimatter and time are intriguing. But what can they tell us about the existence of the universe itself?

The dawn of time

Fourteen billion years ago, the explosion of the Big Bang saw the universe expand from a point of extremely high density and pressure. All the material of the universe came into existence at this point. It is believed that matter and antimatter were created in perfect balance.

In our discussion of the arrow of time in Chapter Five, it was explained how the laws of physics work the same in the backward time direction as in the forward time direction: there is perfect symmetry along the time axis. If we are considering antimatter to be the time-reversed equivalent of normal matter then it is clear why matter and antimatter were created in perfect balance: there is a perfect symmetry in time.

However, we know that if matter and antimatter come into contact then they annihilate each other and all that is left is radiation. So why didn't the matter and antimatter present in the earliest moments in the life of the universe simply annihilate each other, just leaving us with a universe made of light?

Well, there was indeed a huge annihilation of matter by antimatter. But it does appear that there was a very slight imbalance in the annihilation process leaving us with a universe with a slight surplus of matter. No one knows why this slight imbalance occurs, but Einstein summed up the process: "For every one billion particles of antimatter there were one billion and one particles of matter. And when the mutual annihilation was complete, one billionth remained — and that is our present universe."

So, to sum up, it is clear that time plays an integral role in the existence of the universe. It appears that our universe needs time in order that a perfect balance can exist between the positive and the negative. We should be very glad that such a balance can exist, because it appears to be the reason that our universe can exist at all.

But we should also be glad that there was the slightest imbalance in the process of antimatter/matter annihilation, for if that imbalance had not been present then there would be no matter in our universe. And we would not exist.

It appears that balance in our universe is vital for its existence, but it is the slight imbalances that make it such an interesting place to live.

FURTHER READING

Introducing Time by Craig Callender and Ralph Edney
An excellent illustrated book about the physics and philosophy of time.

Simply Einstein by Richard Wolfson
The best introduction to relativity.

Time Machines by Paul J. Nahin
A comprehensive compendium of time machines in science fiction, which also includes a surprisingly strong technical section.

The Feynman Lectures on Physics by Richard Feynman
http://www.feynmanlectures.caltech.edu/
Released online for the first time. Volume One is relevant to the material this book.

The Reference Frame by Luboš Motl
http://motls.blogspot.com
High-level analysis of the latest physics developments written by a divisive character.

PICTURE CREDITS

All photographs are public domain unless otherwise stated.

Hafele and Keating photograph is an historic image provided by Wikipedia.

Bern Clock Tower photograph is courtesy of Mike Lehmann and is provided by Wikimedia Commons.

Einstein optical illusion based on an original idea by Akiyoshi Kitaoka.

Photograph of Drop Tower Bremen is courtesy of ZARM, University of Bremen.

NOTES

[1] The period of the swing is actually only accurate if the amplitude of the swing is not too large. The shorter the swing, the greater the accuracy of the clock.

[2] Ben Majoy, *Redefining Time*, http://www.vice.com/read/time-redefined

[3] http://en.wikipedia.org/wiki/Rietdijk-Putnam_argument As the talk page states: "The Andromeda Paradox is a form of the argument in which two people at the same place and instant have two different universes attached to them."

[4] In his 1913 novel *The World Set Free*, H.G. Wells also uncannily anticipated the development of nuclear power and nuclear weapons, a novel which even influenced the American atomic bomb programme.

[5] James Hartle, *The Physics of 'Now'*, http://arxiv.org/abs/gr-qc/0403001

[6] George Johnson, *Unearthing Prehistoric Tumors, and Debate*, New York Times, December 27th 2010.

[7] Note that, contrary to what many popular science books state, it is not necessary to use general relativity in this situation involving acceleration. Special relativity is sufficient. General relativity would only need to be used if the spacetime was curved.

[8] Frank Tipler, *Rotating Cylinders and the Possibility of Global Causality Violation*, http://tinyurl.com/tiplercylinder

[9] Actually, a rotating object is not travelling at constant velocity, but is continuously accelerating towards the centre of rotation. The Earth's rotation has the effect of slightly reducing the pull of gravity, but this was undetectable to the early astronomers.

[10] John Woodward, *What is the Cause of Inertia?*, http://tinyurl.com/inertiapaper

[11] Carl Brannen, *The Proper Time Geometry*, http://brannenworks.com/a_ptg.pdf

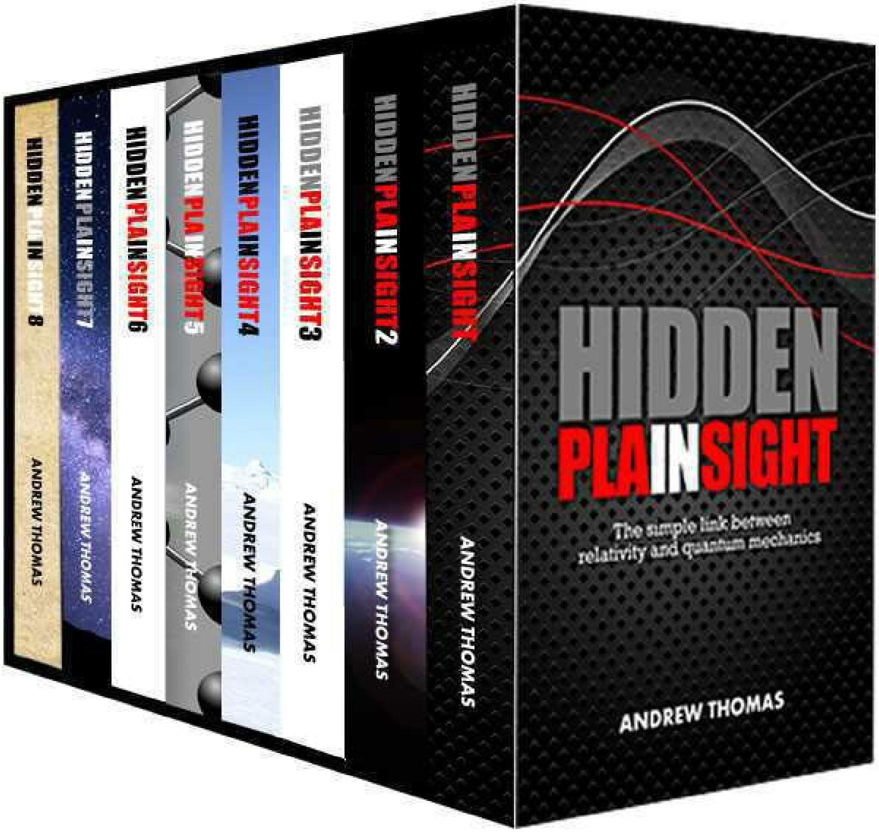
[12] Richard Feynman, *The Character of Physical Law*

[13] Matt Strassler, *The Energy That Holds Things Together*, http://tinyurl.com/energywebsite

[14] The full symmetry is called *CPT symmetry*, meaning that the situation will be changed if the charge (C), parity (P), and time (T) are all reversed.

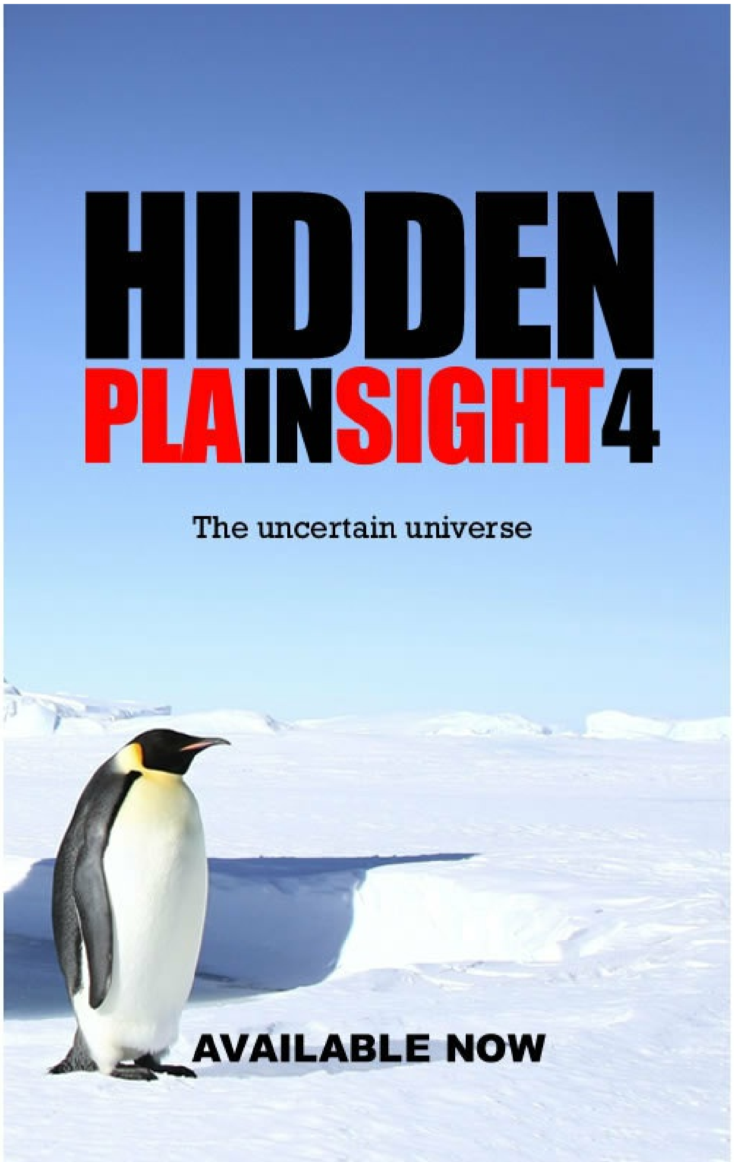
[15] Hilary Putnam, *It Ain't Necessarily So*,

http://tinyurl.com/putnamtimetravel



The "Hidden in Plain Sight" paperbacks

Available now on Amazon


HIDDEN
PLAINSIGHT4
The uncertain universe
AVAILABLE NOW