THE

WORLD

ACCORDING

TO

PHYSICS

JIM

AL-KHALILI

"Clearly written and inviting, even to readers who may not at first consider themselves scientific, *The World According to Physics* is a book that should be read and appreciated by many."

—Jocelyn Bell Burnell, University of Oxford

"A clear, simple, and fascinating account of what physics tells us about our universe, and—crucially—what evidence supports that view, from one of the most talented, inspiring, and informative popularizers of science. A triumph!"

—Ian Stewart, author of *Do Dice Play God?*

"This book presents a comprehensive account of modern physics, weaving together a tapestry of new and familiar topics. Al-Khalili has a distinctively light voice that comes across throughout and works incredibly well."

—Pedro G. Ferreira, author of *The Perfect Theory: A Century of Geniuses and the Battle over General Relativity*

THE WORLD ACCORDING TO PHYSICS

THE WORLD ACCORDING TO PHYSICS

JIM AL-KHALILI

CONTENTS

This book is an ode to physics.

I first fell in love with physics when I was a teenager. Admittedly, this was partly because I realised I was good at it. The subject seemed to be a fun mix of puzzle-solving and common sense, and I enjoyed playing with the equations, manipulating the algebraic symbols, and plugging in numbers so that they revealed the secrets of nature. But I also realised that if I wanted satisfying answers to the many deep questions about the nature of the universe and the meaning of existence bubbling up in my teenage mind, then physics was the subject I had to study. I wanted to know: What are we made of? Where do we come from? Does the universe have a beginning, or an end? Is it finite in extent, or does it stretch out to infinity? What was this thing called quantum mechanics that my father had mentioned to me? What is the nature of time? My quest to find

answers to these questions has led to a life spent studying physics. I have some answers to my questions now; others I am still searching for.

Some people turn to religion or some other ideology or belief system to find answers to life's mysteries. But for me, there is no substitute for the careful hypothesising, testing, and deducing of facts about the world that are the hallmark of the scientific method. The understanding we have gained through science—and physics in particular—of how the world is made up and how it works is, in my view, not just one of many equally valid ways of reaching the 'truth' about reality. It is the *only* reliable way we have.

No doubt many people never fell in love with physics, as I did. Perhaps they were turned off from studying science because they decided, or perhaps were told by others, that it is a hard—or a geeky—subject. And to be sure, getting to grips with the subtleties of quantum mechanics can bring on a headache. But the wonders of our universe can and should be appreciated by everyone, and gaining a basic understanding doesn't take a lifetime of study. In this book, I

want to describe why physics is so wonderful, why it is such a fundamental science, and why it is so crucial to our understanding of the world. The grand scope and sweep of physics today are breathtaking. That we now know what (almost) everything we see in the world is made of and how it holds together; that we can trace back the evolution of the entire universe to fractions of a second after the birth of space and time themselves; that through our knowledge of the physical laws of nature we have developed, and continue to develop, technologies that have transformed our lives—this is all pretty staggering. I still find myself thinking, as I write this: How can anyone *not* love physics?

This book is intended to serve as an introduction to some of the most profound and fundamental ideas in physics. But the topics I cover are not ones you will likely have encountered at school. For some readers, the book may be a first invitation into physics—one that will entice you to learn more about it, maybe even pursue it as a lifelong journey of study and discovery, as I have. To others, who may have gotten off on the

wrong foot with physics early on, it may serve as a gentle reintroduction. For many, it may provoke wonder at just how far humanity has come in its quest to understand.

To convey a working knowledge of what physics tells us about the nature of our world, I have selected an array of the most important concepts in modern physics and attempted to show how they link together. We'll survey the vast range of this conceptual landscape, from the physics of the largest cosmic scales to that of the smallest quantum level; from physicists' quest to unify the laws of nature to their search for the simplest possible physical principles governing life; from the speculative frontiers of theoretical research to the physics that underpins our everyday experiences and technologies. I will also offer readers some new perspectives: ideas that we physicists have learnt to accept, but which we haven't done a very good job of conveying to those outside our innermost circles of experts. For example, down at the subatomic scale, separated particles communicate with each other instantaneously despite being far apart, in a way that violates common

sense. This property, called nonlocality, may force us ultimately to revise our entire understanding of the structure of space itself. But, sadly, many non-physicists—and indeed some physicists— misunderstand or misinterpret what this really means.

A criticism levelled (usually by theoretical physicists) at many popular science books covering fundamental concepts in physics is that they don't always help the lay reader grasp what these concepts actually mean. In my view, this is because the physicists who truly understand the concepts, and who write the research papers and come up with new theories, are not necessarily the best at explaining their own ideas to non-physicists. But, in turn, those who may have more experience and success with communicating their work to the public may not understand certain concepts deeply enough themselves to go beyond simple analogies. Even if one understands the physics and can successfully (I hope) communicate with non-physicists, it is not a small challenge to explain terms like gauge invariance, duality, eternal inflation, the holographic

principle, conformal field theories, anti-de Sitter spaces or vacuum energy in a way that conveys real insight into the physics involved, without involving complex mathematics. I have done my best, but there may well be some readers who feel I could have done better. And, of course, this will be true.

Nevertheless, if you wish to delve more deeply into any particular topics which I only touch on briefly here, then there are many books that do this brilliantly. I list at the end of the book some of those I believe you would find the most accessible and enlightening. Many of the books on this list describe the journey of scientific progress—how physics has developed over the millennia since the ancient Greeks, how discoveries were made, and how theories and hypotheses were proposed and discarded. These books often focus on the revolutions that have overturned previously held views about the universe and describe the leading players in these historical accounts. In this short book, however, I won't look back on how far we've come; nor will I say too much about how far we have yet to go (since I

don't know, and also because I suspect it is still a long way), although I will focus in chapter 8 on what we *know* we don't know.

I have no particular theory to plug, either. For example, when it comes to reconciling quantum mechanics with general relativity (the holy grail of modern theoretical physics), I do not subscribe to either of the two main camps working towards this goal: I am neither a string theory advocate nor a loop quantum gravity fan,[1] since neither theory falls within my particular specialism; and when it comes to interpreting the meaning of quantum mechanics, I am neither a 'Copenhagenist' nor a 'many worlds' enthusiast.[2] But, this won't stop me from being somewhat polemical about these issues now and then.

I will also try not to become too embroiled in philosophical or metaphysical musings, even though there is a temptation to do so when one is discussing some of the more profound ideas at the forefront of physics, whether on the nature

---

1 I will of course explain what these ideas involve later.

2 Again, I will explain later.

of space and time, the various interpretations of quantum mechanics, or even the meaning of reality itself. I do not mean by this that physics does not need philosophy. To give you an idea of how philosophy feeds into my subject at the most fundamental level, you may be surprised to know that physicists cannot yet even agree on whether the job of physics is to figure out how the world *really* is, as Einstein believed—to reach some ultimate truth that is waiting out there to be discovered—or whether it is to build models of the world and to come up with our best current stab at *what we can say* about reality, a reality that we may never truly know. On this matter, I am on the side of Einstein.

To put it simply, I would argue that physics gives us the tools to understand the entire universe. The study of physics is a search for explanations, but to embark on that search we must first ask the right questions, something philosophers are very good at.

And so, we will begin our journey in a suitably humble frame of mind, one that, if we're honest, we all share—as children, as adults, and with

generations past and future: one of not know-ing. By thinking about what we don't yet know, we can think about how we can best find out. It is the many questions we have asked over the course of our human history that have given us an ever-more-accurate picture of the world we know and love.

So, here is the world according to physics.

THE WORLD ACCORDING TO PHYSICS

THE AWE OF UNDERSTANDING

While stories will always be a vital part of human culture, even in science—and our lives would be the poorer without them—modern science has now replaced many of the ancient mythologies and accompanying superstitious beliefs. A good example of how we have demystified our approach to understanding the world is the creation myths. Since the dawn of history, humankind has invented stories about the origins of our world, and deities that were instrumental in its creation, from the Sumerian god Anu, or Sky Father, to the Greek myths about Gaia being created out of Chaos and the Genesis myths of the Abrahamic religions, which are still believed as literal truths in many societies around the world. It may appear to many non-scientists

that our modern cosmological theories about the origins of the universe are themselves no better than the religious mythologies they replace—and, if you look at some of the more speculative ideas in modern theoretical physics, you might agree that those who feel this way have a point. But through rational analysis and careful observation—a painstaking process of testing and building up scientific evidence, rather than accepting stories and explanations with blind faith—we can now claim with a high degree of confidence that we know quite a lot about our universe. We can also now say with confidence that what mysteries remain need not be attributed to the supernatural. They are phenomena we have yet to understand—and which we hopefully will understand one day through reason, rational enquiry, and, yes . . . physics.

Contrary to what some people might argue, the scientific method is *not* just another way of looking at the world, nor is it just another cultural ideology or belief system. It is the way we learn about nature through trial and error, through experimentation and observation,

through being prepared to replace ideas that turn out to be wrong or incomplete with better ones, and through seeing patterns in nature and beauty in the mathematical equations that describe these patterns. All the while we deepen our understanding and get closer to that 'truth'— the way the world *really* is.

There can be no denying that scientists have the same dreams and prejudices as everyone else, and they hold views that may not always be entirely objective. What one group of scientists calls 'consensus', others see as 'dogma'. What one generation regards as established fact, the next generation shows to be naïve misunderstanding. Just as in religion, politics, or sport, arguments have always raged in science. There is often a danger that, all the while a scientific issue remains unresolved, or at least open to reasonable doubt, the positions held by each side of the argument can become entrenched ideologies. Each viewpoint can be nuanced and complex, and its advocates can be just as unshakable as they would be in any other ideological debate. And just as with societal attitudes on religion, politics, culture, race,

or gender, we sometimes need a new generation to come along, shake off the shackles of the past, and move the debate forward.

But there is also a crucial distinction to science, when compared with other disciplines. A single careful observation or experimental result can render a widely held scientific view or long-standing theory obsolete and replace it with a new worldview. This means that those theories and explanations of natural phenomena that have survived the test of time are the ones we trust the most; they are the ones we are most confident about. The Earth goes around the Sun, not the other way around; the universe is expanding, not static; the speed of light in a vacuum always measures the same no matter how fast the measurer of that speed is moving; and so on. When a new and important scientific discovery is made, which changes the way we see the world, not all scientists will buy into it immediately, but that's *their* problem; scientific progress is inexorable, which, by the way, is *always* a good thing: knowledge and enlightenment are always better than ignorance. We start with not knowing, but we

seek to find out . . . and, though we may argue along the way, we cannot ignore what we find. When it comes to our scientific understanding of how the world is, the notion that 'ignorance is bliss' is a load of rubbish. As Douglas Adams once put it: 'I'd take the awe of understanding over the awe of ignorance any day.'[1]

WHAT WE DON'T KNOW

It is also true that we are constantly discovering how much more there is that we don't yet know. Our growing understanding yields a growing understanding of our ignorance! In some ways, as I will explain, this is the situation we have in physics right now. We are currently at a moment in history when many physicists see, if not a crisis in the subject, then at least the building up of a head of steam. It feels as though something has to give. A few decades ago, prominent physicists such as Stephen Hawking were asking, 'Is the end in sight

1 Douglas Adams, *The Salmon of Doubt: Hitchhiking the Galaxy One Last Time* (New York: Harmony, 2002), 99.

for theoretical physics?'[2] with a 'theory of everything' potentially just around the corner. They said it was just a matter of dotting the 'i's and crossing the 't's. But they were wrong, and not for the first time. Physicists had expressed similar sentiments towards the end of the nineteenth century; then along came an explosion of new discoveries (the electron, radioactivity, and X-rays) that couldn't be explained by the physics known at the time and which ushered in the birth of modern physics. Many physicists today feel that we might potentially be on the verge of another revolution in physics as big as that seen a century ago with the birth of relativity and quantum mechanics. I am not suggesting that we are about to discover some fundamental new phenomenon, like X-rays or radioactivity, but there may yet be a need for another Einstein to break the current deadlock.

The Large Hadron Collider has not yet followed up on its 2012 success in detecting the Higgs boson, and thereby confirming the ex-

---

2 This was the title of an article Hawking wrote in 1981: S. W. Hawking, *Physics Bulletin* **32**, no. 1 (1981): 15–17.

istence of the Higgs field (which I will discuss later); many physicists were hoping for the discovery of other new particles by now, which would help resolve long-standing mysteries. And we still don't understand the nature of the dark matter holding galaxies together or the dark energy that is ripping the universe apart; nor do we have answers to fundamental questions like why there is more matter than antimatter; why the properties of the universe are so finely tuned to allow for stars and planets, and life, to exist; whether there is a multiverse; or whether there was anything before the Big Bang that created the universe we see. There is still so much left that we cannot explain. And yet, it is hard *not* to be dazzled by our success so far. While some scientific theories may turn out to be connected to each other at a deeper level than we thought, and others may turn out to be entirely wrong, no one can deny just how far we've come.

Sometimes, in the light of new empirical evidence, we realise that we were barking up the wrong tree. Other times we simply refine an idea that turns out not to be wrong, but just a rough

approximation that we improve upon to gain a more accurate picture of reality. There are some areas of fundamental physics that we might not be entirely happy with, where we know deep down that we've not heard the final word, but which we nevertheless continue to rely on for the time being because they are useful. A good example of this is Newton's universal law of gravitation. It is still referred to, grandly, as a 'law' because scientists at the time were so confident that it was the last word on the subject that they elevated its status above that of a mere 'theory'. The name stuck, despite the fact that we now know their confidence was misplaced. Einstein's general theory (note that it's called a theory) of relativity replaced Newton's law, because it gives us a deeper and more accurate explanation of gravity. And yet, we still use Newton's equations to calculate the flight trajectories of space missions. The predictions of Newtonian mechanics may not be as accurate as those of Einstein's relativity, but they are still good enough for nearly all everyday purposes.

Another example that we are still working on is the Standard Model of particle physics. This is

an amalgamation of two separate mathematical theories, called electroweak theory and quantum chromodynamics, which together describe the properties of all the known elementary particles and the forces acting between them. Some physicists think of the Standard Model as nothing more than a stopgap until a more accurate and unified theory is discovered. And yet, it is remarkable that, as it stands now, the Standard Model can tell us everything we need to know about the nature of matter: how and why electrons arrange themselves around atomic nuclei, how atoms interact to form molecules, how those molecules fit together to make up everything around us, how matter interacts with light (and therefore how almost all phenomena can be explained). Just one aspect of it, quantum electrodynamics, underpins all of chemistry at the deepest level.

But the Standard Model cannot be the final word on the nature of matter, because it doesn't include gravity and it doesn't explain dark matter or dark energy, which between them make up most of the stuff of the universe. Answering some questions naturally leads to others, and

physicists continue their search for physics 'beyond the Standard Model' in an attempt to address these lingering but crucial unknowns.

HOW WE PROGRESS

More than any other scientific discipline, physics progresses via the continual interplay between theory and experiment. Theories only survive the test of time as long as their predictions continue to be verified by experiments. A good theory is one that makes new predictions that can be tested in the lab, but if those experimental results conflict with the theory, then it has to be modified, or even discarded. Conversely, laboratory experiments can point to unexplained phenomena that require new theoretical developments. In no other science do we see such a beautiful partnership. Theorems in pure mathematics are proven with logic, deduction, and the use of axiomatic truths. They do not require validation in the real world. In contrast, geology, ethology or behavioural psychology are mostly observational sciences in which advances in our

understanding are made through the painstaking collection of data from the natural world, or via carefully designed laboratory tests. But physics can *only* progress when theory and experiment work hand in hand, each pulling the other up and pointing to the next foothold up the cliffside.

Shining a light on the unknown is another good metaphor for how physicists develop their theories and models, and how they design their experiments to test some aspect of how the world works. When it comes to looking for new ideas in physics, there are, very broadly, two kinds of researchers. Imagine you're walking home on a dark, moonless night when you realise that there's a hole in your coat pocket through which your keys must have fallen at some point along your route. You know they have to be somewhere on the ground along the stretch of pavement you've just walked, so you retrace your steps. But do you only search the patches bathed in light beneath lampposts? After all, while these areas cover only a fraction of the pavement, at least you will see your keys if they are there. Or do you grope around in the dark

stretches in between the pools of lamplight? Your keys may be more likely to be here, but they will also be more difficult to find.

Similarly, there are lamppost physicists and searchers in the dark. The former play it safe and develop theories that can be tested against experiment—they look where they can see. This means they tend to be less ambitious in coming up with original ideas, but they achieve a higher success rate in advancing our knowledge, albeit incrementally: evolution, not revolution. In contrast, the searchers in the dark are those who come up with highly original and speculative ideas that are not so easy to test. Their chances of success are lower, but the payoff can be greater if they are right, and their discoveries can lead to paradigm shifts in our understanding. This distinction is far more prevalent in physics than in other sciences.

I have sympathy for those who get frustrated by the searchers and the dreamers, who often work in esoteric areas like cosmology and string theory, for these are the people who think nothing of adding a few new dimensions here or there if it makes their maths prettier, or to hypothesise

an infinity of parallel universes if it reduces the strangeness in ours. But there have been some famous examples of searchers who have struck gold. The twentieth-century genius Paul Dirac was a man driven by the beauty of his equations, which led him to postulate the existence of antimatter several years before it was discovered in 1932. Then there's Murray Gell-Mann and George Zweig, who in the mid-1960s independently predicted the existence of quarks when there was no experimental evidence to suggest such particles existed. Peter Higgs had to wait half a century for his boson to be discovered and the theory that bears his name to be confirmed. Even the quantum pioneer Erwin Schrödinger came up with his eponymous equation with nothing more than inspired guesswork. He picked the right mathematical form of equation even though he didn't yet know what its solution meant.

What unique talents did all these physicists have? Was it intuition? Was it a sixth sense that allowed them to sniff out nature's secrets? Possibly. The Nobel Prize winner Steven Weinberg believes it is the aesthetic beauty in the mathematics that

has guided great theoreticians like Paul Dirac and the great nineteenth-century Scottish physicist James Clerk Maxwell.

But it is also true that none of these physicists worked in isolation, and their ideas still had to be consistent with all established facts and experimental observations.

THE SEARCH FOR SIMPLICITY

The true beauty of physics, for me, is found not only in abstract equations or in surprising experimental results, but in the deep underlying principles that govern the way the world is. This is a beauty that is no less awe-inspiring than a breathtaking sunset or a great work of art such as a Leonardo da Vinci painting or Mozart sonata. It is a beauty that lies not in the surprising profundity of the laws of nature, but in the deceptively simple underlying explanations (where we have them) for where those laws come from.[3]

---

3 Of course, beauty need not only be associated with simplicity. Just as with great art or music, there can also be beauty in the sheer complexity of some physical phenomena.

A perfect example of the search for simplicity is science's long and continuing journey to discover the basic building blocks of matter. Take a look around you. Consider the sheer range of materials that make up our everyday world: concrete, glass, metals, plastics, wood, fabrics, foodstuffs, paper, chemicals, plants, cats, people . . . millions of different substances, each with its own distinctive properties: squidgy, hard, runny, shiny, bendy, warm, cold. . . . If you knew nothing of physics or chemistry, you might imagine that most materials have little in common with each other; and yet we know that everything is made of atoms, and that there is only a finite number of different kinds of atoms.

But our quest for ever-deeper simplicity does not stop there. Thinking about the structure of matter goes all the way back to the fifth century BC in ancient Greece, when Empedocles first proposed that all matter consisted of four fundamental 'elements' (his 'fourfold roots of everything'): earth, water, air, and fire. In contrast to this simple idea, and around the same time, two other philosophers, Leucippus and his

pupil Democritus, proposed that all matter was composed of tiny indivisible 'atoms'. However, these two promising ideas conflicted with each other. While Democritus believed that matter was ultimately made of fundamental building blocks, he thought there would be an infinite variety of such different atoms; whereas Empedocles, who proposed that everything was ultimately made up of just four elements, argued that these elements were continuous and infinitely divisible. Both Plato and Aristotle promoted the latter theory and rejected Democritus's atomism, believing that its simplistic mechanistic materialism could not produce the rich diversity of beauty and form of the world.

What the Greek philosophers were doing was not true science as we understand it today—apart from a few notable exceptions, such as Aristotle (the observer) and Archimedes (the experimenter), their theories were often not much more than idealised philosophical concepts. Nevertheless, today, through the tools of modern science, we know that both of those ancient ideas (atomism and the four elements) were, in spirit at least,

along the right lines: that all the stuff making up our world, including our own bodies, and including everything we see out in space—the Sun, the Moon, and the stars—is all made of fewer than a hundred different types of atoms. We also now know that atoms have internal structure. They are made of tiny, dense nuclei surrounded by clouds of electrons while the nucleus itself is made up of smaller constituents: protons and neutrons, which are in turn made of even more fundamental building blocks called quarks.

So, despite the apparent complexity of matter and the immeasurable variety of substances that can be made up from the chemical elements, the truth is that the ancients' quest for simplicity didn't go far enough. As we understand physics today, all the matter we see in the world is made up of not the four classical elements of the Greeks, but just three elementary particles: the 'up' quark, the 'down' quark, and the electron. That's it. Everything else is just detail.

And yet the job of physics is more than just classifying what the world is made of. It is about finding the correct explanations for the natural

phenomena we observe and the underlying principles and mechanisms that account for them. While the ancient Greeks might have debated passionately about the reality of atoms or the abstract connection between 'matter' and 'form', they had no idea how to explain earthquakes or lightning, let alone astronomical events such as the phases of the Moon or the occasional appearance of comets—although this didn't prevent them from trying.

We have come a very long way since the Greeks of antiquity, and yet there is also plenty that we still have to understand and explain. The physics I will cover in this book is mostly the stuff we are confident about. Throughout, I will explain *why* we are confident and point out what is speculative and where there may be some wiggle room. Naturally, I anticipate that some parts of the story will become out-of-date in the future. Indeed, an important discovery might be made the day after this book's publication that revises some aspect of our understanding. But that is the nature of science. *Mostly*, what you will read about in this book is

established beyond reasonable doubt to be the way the world *is*.

In the next chapter, I explore the idea of scale. No other science so brazenly addresses such a vast range of scales, of time, space, and energies, as physics does, from the unimaginably tiny quantum world to the entire cosmos, and from the blink of an eye to eternity.

After gaining an appreciation for the scope of what physics can explain, we will begin on our journey in earnest, starting with the three 'pillars' of modern physics: relativity, quantum mechanics, and thermodynamics. In order to paint the picture of our world that physics has given us, we must first prepare the canvas, and in this case the canvas is space and time. Everything that happens in the universe comes down to events that take place somewhere in space and at some moment in time. And yet, we will see in chapter 3 that we cannot separate the canvas from the painting. Space and time themselves are an integral part of reality. You may be shocked to discover just how different the physicist's view of space and time is from our everyday,

commonsense one, for it relies on Einstein's general theory of relativity, which describes the nature of space and time and defines how we think about the fabric of the cosmos. Once this canvas is ready, we can proceed to prepare our paints. In chapter 4, I define what a physicist means by matter and energy, the stuff of the universe: what it consists of, how it was created, and how it behaves. One can think of this chapter as a companion to the previous one, because I also describe how matter and energy are intimately related to the space and time in which they exist.

In chapter 5, I plunge into the world of the very small, zooming in and shrinking down to study the nature of the fundamental building blocks of matter. This is the quantum world, our second pillar of modern physics, where matter behaves very differently from our everyday experiences, and where our grip on what is real becomes increasingly tenuous. And yet . . . our understanding of the quantum is far more than a flight of fancy or mere intellectual diversion; without an understanding of the rules govern-

ing the building blocks of matter and energy, we would not have been able to build our modern technological world.

In chapter 6, we zoom out of the quantum world to see what happens when we put many particles together to make up larger, more complex systems. What do physicists mean by order, disorder, complexity, entropy, and chaos? Here, we encounter the third pillar of physics, thermodynamics—the study of heat, energy, and the properties of matter in bulk. We are led inevitably to ask what makes life itself so special. How is living matter so different from non-living matter? After all, life must be subject to the same laws of physics as everything else. In other words, can physics help us understand the difference between chemistry and biology?

In chapter 7, I explore one of the most profound ideas in physics, the notion of unification: the way we have sought, and found, over and over again, universal laws that bring together seemingly disparate phenomena in nature under one unifying description or theory. I conclude the chapter with a look at some of the

front-runners for an all-encompassing physical 'theory of everything'.

By chapter 8 we will have reached the limit of what we currently understand about the physical universe and can finally dip our toes in the vast ocean of the unknown. I explore some of the mysteries we are currently struggling with and speculate upon whether we are close to solving them.

In the penultimate chapter, I discuss how the interplay of theory and experiment in physics has led to the technologies on which our modern world is built. For example, without quantum mechanics, we would not have been able to understand the behaviour of semiconductors or invent the silicon chip, on which all of modern electronics is founded, and I would not be typing these words on my laptop. I will also take a look into the future and predict how current research into quantum technologies is going to revolutionise our world in unimaginable ways.

In the final chapter, I explore the notion of scientific truth, particularly in a 'post-truth' society in which many people remain suspicious

of science. How does the process of science differ from other human activities? Is there such a thing as absolute scientific truth? And if the job of science is to seek out deep truths about nature, how should scientists convince wider society of the value of the scientific enterprise: the forming and testing of hypotheses, and rejecting them if they do not fit the data? Will science ever come to an end one day when we know all there is to know? Or will the search for answers continue to lead us deeper down an ever-expanding abyss?

I promised you in the preface that I would try not to get too tangled up in philosophical musings, and yet here I am doing just that, and this is still only the Introduction. So, I will take a deep breath and start us off again, gently, with a sense of scale.

CHAPTER 2

SCALE

Unlike philosophy, logic, or pure mathematics, physics is both an empirical and a quantitative science.[1] It relies on the testing and verification of ideas through reproducible observation, measurement, and experimentation. While physicists can sometimes propose exotic or outlandish mathematical theories, the only true measure of their efficacy and power is whether they describe phenomena in the real world against which we can test them. This is why Stephen Hawking never won a Nobel Prize for his work in the mid-1970s on the way black holes radiate energy, a phenomenon known as Hawking radiation: the Nobel is only awarded to theories or discover-

---

1 Just for completeness, I should add that during the past couple of decades a new discipline called experimental philosophy has emerged.

ies that have been confirmed experimentally. Likewise, Peter Higgs and others who made a similar prediction had to wait half a century for the existence of the Higgs boson to be confirmed at the Large Hadron Collider.

It is also the reason why physics as a scientific discipline only began to make truly impressive advances once the tools and instruments necessary to test theories—through observation, experimentation, and quantitative measurement—had been invented. The ancient Greeks may have been brilliant at abstract thinking, developing subjects such as philosophy and geometry to a level of sophistication that is still valid today, but—Archimedes aside—they were not particularly famous for their experimental prowess. The world of physics only really came of age in the seventeenth century, thanks to a large extent to the invention of the two most important instruments in all of science: the telescope and the microscope.

If we were only able to understand the world we can see with our naked eyes, then physics would not have got very far. The range of wavelengths

that can be 'seen' by the human eye is just a sliver of the full electromagnetic spectrum, and our eyes are constrained to discerning only those objects that are not too small and not too far away. While we can, in principle, see out to infinity, provided a sufficient number of photons make it to our eyes (and given an infinite amount of time for them to reach us!), this would not likely provide us with much useful detail. But, once the microscope and the telescope were invented, they opened up windows on the world that dramatically increased our understanding, magnifying the very small and bringing closer the very far away. At last, we could make observations, and detailed measurements, to test and refine our ideas.

On the 7th of January 1610, Galileo pointed his modified and improved spyglass up towards the heavens and banished forever the notion that we were at the centre of the cosmos.[2] He observed

---

2 No doubt historians of science will dispute this simplistic claim. Galileo did not suddenly establish heliocentrism with his observations and really only offered suggestive facts (like Jupiter's moons).

four of the moons of Jupiter and correctly inferred that Copernicus's heliocentric model was correct—that the Earth goes around the Sun and not vice versa. By observing bodies in orbit around Jupiter, he showed that not all celestial bodies revolve around us. The Earth isn't at the centre of the cosmos, but is just another planet, like Jupiter, Venus, and Mars, orbiting the Sun. With that discovery, Galileo ushered in modern astronomy.

But it wasn't just a revolution in astronomy that Galileo would bring about. He also helped put the scientific method itself on a firmer foundation. Building on the work of the medieval Arab physicist Ibn al-Haytham, Galileo 'mathematised' physics itself. In developing mathematical relationships that describe, and indeed predict, the motions of bodies, he showed beyond doubt that, as he put it, the book of nature 'is written in mathematical language.'[3]

---

3 A quote from Galileo's famous book, *The Assayer* (Italian: *Il Saggiatore*), published in Rome in 1623.

At the opposite end of the scale to Galileo's astronomical observations, a very different new world was opened up by Robert Hooke and Antonie van Leeuwenhoek with the microscope. Hooke's famous book, *Micrographia,* published in 1665, contains stunning drawings of miniature worlds, from the eye of a fly and the hairs on the back of a flea to individual cells in plants, that no one had ever witnessed before.

Today, the range of scales open to exploration by humankind is astounding. With electron microscopes we can see individual atoms, just a tenth of a millionth of a millimetre across, and with giant telescopes we can gaze out to the furthest reaches of the observable universe 46.5 billion light-years away.[4] No other science studies such a span in scale. In fact, forget resolutions to the size of atoms, a team at the

---

4 The most distant light we can see, from the edge of the observable universe, has been travelling towards us for over 13 billion years and so shows us what the universe was like when it was very young. However, due to the expansion of space, the origin of this light is now much further away than 13 billion light-years.

University of St Andrews in Scotland recently showed me something mind-blowingly impressive when it comes to measuring the smallest length scales. They've come up with a way of measuring the wavelength of light using a device called a wavemeter, to an accuracy of a single attometre—or one-thousandth of the diameter of a proton. They did this by passing laser light through a short optical fibre, which scrambles the light into a grainy pattern called 'speckle', and then tracking how this pattern changes with the minutest of adjustments to the wavelength of the light.

And it is not only the range of *length* scales that physics embraces; we can also measure time from the minutest fractions of the twinkling of an eye to cosmic eternities. Here is a stunning example. In an experiment carried out in Germany in 2016, physicists measured a period of time almost too short to imagine. They were studying a phenomenon called the photoelectric effect, in which photons free up electrons by knocking them out of atoms. The process was first explained correctly by Einstein

in 1905 in a famous paper for which he won the Nobel Prize many years later (and not for his work on the theories of relativity, as you may have thought). Today, this process of knocking electrons out of materials is called photoemission and is the way we turn sunlight into electricity in solar cells.

In the 2016 experiment, two special lasers were used. The first fired an almost unimaginably short pulse of ultraviolet laser light at a jet of helium gas. The duration of this pulse was a mere ten thousandth of a trillionth of a second, or 100 attoseconds (10^{-18} seconds).[5] The second laser was lower in energy (its frequency being in the infrared range) and its pulse duration was a little longer than the first. Its job was to capture the escaping electrons, allowing the researchers to calculate how long it had taken them to be knocked out. The researchers found that this was even quicker: a mere tenth of the duration of the first laser pulse. What is inter-

---

5 There are more attoseconds in a single second than there have been seconds since the Big Bang.

esting about this result is that the knocked-out electrons actually drag their heels a little. You see, helium atoms each contain two electrons, and the ones that are knocked out feel the influence of the partner they leave behind, which, ever so slightly, delays the ejection process. It is staggering to think that a physical process taking just a few attoseconds can actually be measured like this in the lab.

In my own field of nuclear physics, there are processes that are even faster than this, although these cannot be measured directly in the lab. Instead, we develop computer models to explain the different structures of atomic nuclei and the processes that take place when two nuclei collide and react. For example, the first step in nuclear fusion—when two heavy nuclei come together like coalescing drops of water to make an even heavier nucleus— involves the very rapid reorganization of all the protons and neutrons from both nuclei into the new combined nucleus. This quantum process takes less than a zeptosecond (10^{-21} seconds).

At the other extreme of the time scale, cosmologists and astronomers have been able to work out the age of (our part of) the universe so precisely that we are now confident that the Big Bang took place 13.8242 billion years ago (give or take a few million years). Our confidence in the accuracy of this value may sound arrogant to some—and even unbelievable to those who still cling to the medieval idea that the universe is only six thousand years old—so let me explain how we come to this figure.

Let me first make two important assumptions, which I will discuss in more detail later on, but will now just say that they are both supported strongly by observational evidence: (1) that the laws of physics are the same everywhere in our universe, and (2) that space looks the same in all directions (the same density and distribution of galaxies). This gives us confidence that we can use the observations we make from Earth, or via satellite observatories in orbit around the Earth, to learn about the entire cosmos. Doing this has allowed us to work out the age of the universe in several different ways.

For example, we can learn a lot by studying the stars in our galaxy. We know how long stars can live, depending on their size and brightness, which determines how fast they burn via thermonuclear fusion. This means we can work out the age of the oldest stars, which sets a lower limit on how old our galaxy is, which in turn gives us a lower limit on the age of the universe. Since the oldest stars are about 12 billion years, the universe cannot be younger than that.

Then, by measuring the brightness and colour of the light entering our telescopes from distant galaxies, we can work out how fast the universe is expanding, both now and in the past. The further out we look, the further back in time we are probing, since the light we see will have taken billions of years to reach us and is thus bringing us information about the distant past. And if we know how fast the universe has been expanding, we can wind back the clock to a time when everything was squeezed together in the same place: the moment of the universe's birth.

Quite separately, by studying the tiny variations in the temperature of deep space (the so-called cosmic microwave background) we can get an accurate snapshot of the universe as it was before any stars and galaxies had even formed, just a few hundred thousand years after the Big Bang. This allows us to pinpoint the age of the universe even more precisely.

While it is one thing to say that physics allows us to learn about the universe at the shortest and longest distance and time scales, what I find equally remarkable is that we have discovered laws of physics that apply across the entirety of these ranges. Maybe you do not find this surprising; maybe it is natural for you to assume that the laws of nature that operate on the human scale should also work on other scales of distance, time, and energy. But this should be far from obvious.

To explore this further, I will introduce three concepts that are not always taught to students of physics, but which most certainly should be: universality, symmetry, and reductionism.

UNIVERSALITY

The first 'universal'[6] law of physics was discovered by Isaac Newton.[7] Whether or not he saw an apple fall from a tree on his mother's farm, triggering him to develop his law of gravitation, or what the mathematical formula articulating this law looks like, are not of importance here. The crucial point is that Newton realised that the force that pulls an apple to the ground has the same origin as the force that keeps the Moon in orbit around the Earth—that a simple mathematical relation can describe both processes equally well. The way objects behave due to gravity here on Earth is the same as the behaviour of the

---

6 I am using the word here in a very general sense and not in the more specific way as understood by some physicists working in the field of statistical mechanics. There, the term 'universality', as introduced by the American physicist Leo Kadanoff in the 1960s, is the observation that there are properties for a class of physical systems that do not depend on their detailed structure and dynamics but can instead be deduced from a few global parameters.

7 In fact, Robert Hooke's work on gravitation preceded that of Newton.

Moon around the Earth, the planets around the Sun, and the Sun around the centre of the Milky Way galaxy. The gravitational force that shapes life on Earth is the same force that has shaped the entire universe since the Big Bang. The fact that Newton's description of gravity was superseded by Einstein's more accurate one more than two centuries later does not detract from this insight about the *universality* of gravity.

Einstein's general theory of relativity, which improved on the predictions of Newton, also gave us an entirely new description of reality, which I will explore in greater depth in the next chapter. Indeed, Einstein's theory demonstrates a universality that is quite astonishing, and I will mention just one aspect of it here to highlight what I mean. The beautiful mathematical construct that Einstein presented to the world in 1915 is also still our current best theory on the nature of space and time, and it is extremely accurate. It also correctly predicts that a gravitational field will *slow down* the passage of time: the stronger the field, the slower time runs. This effect has the strange consequence that time

ticks by ever so slightly slower in the Earth's core (deep within its gravitational well) than it does on the surface. This difference in age that has accumulated over the four and half billion years of our planet's existence means that the core is in fact two and a half years younger than the crust. Put another way, for every sixty years of Earth's history, its core has aged one second less than its crust. This figure has been calculated using the formula from general relativity, and it is not at all obvious how we might go about checking it experimentally, but such is our trust in the formula that no physicist is in any real doubt about its veracity.

If you think about the above prediction, you might find it somewhat paradoxical. After all, if we were to drill a hole through the Earth, then travel down to its centre, we would no longer feel the effects of gravity since the Earth would be pulling us equally in all directions—we would feel weightless. However, the effect on time is not due to the strength of the gravitational *force* at the centre of the Earth, which is zero, but rather to the gravitational '*potential*' there. This is the amount of energy needed to pull a body from that

location out to a place far from Earth's gravity entirely. A physicist would say that the core of the Earth is in the deepest part of the Earth's potential well, where the slowing of time is greatest.

We can even measure the difference in the rate of flow of time over a height of just a few metres. A clock upstairs in your house is in a slightly weaker gravitational potential (further from the Earth's core) than a clock downstairs, and so runs ever so slightly faster. But this effect is extremely tiny: the two clocks would be out of sync by just one second every hundred million years.

If you are feeling sceptical about all this, let me assure you that the quantitative effect of gravity on time is very real indeed; if we didn't take it into account in modern telecommunications, the smartphone in your pocket would not be able to pinpoint your location anywhere nearly as accurately. Where you are on Earth relies on your phone sending and receiving signals from several GPS satellites in orbit. The time it takes for these electromagnetic waves to cover the distance has to be known to within just a few hundredths of a microsecond (so that your location

can be pinpointed to within a few metres). But this doesn't work if we assume that time runs at the same rate everywhere. In fact, the highly precise atomic clocks on board satellites gain around 40 millionths of a second each day and so must be deliberately slowed down in order to match the rate of slower Earth-bound clocks. Without this, satellite clocks would gain time and your GPS location would drift by over ten kilometres each day—rendering the information useless.

What is also remarkable is that those same equations of general relativity that predict the way gravity causes tiny modifications to the rate at which clocks tick can also tell us about the longest time scales imaginable, mapping the history of the universe over billions of years all the way back to the Big Bang, and even predicting its future. Einstein's theory of relativity applies equally well at the shortest and longest intervals of time.

But this universality only stretches so far. We know that, at the very tiniest of length and time scales, the physics of our everyday world (whether according to Newton or Einstein) breaks down and must be replaced by the predictions of quantum

mechanics. Indeed, as I will explain in the coming chapters, the very definition of time according to quantum theory differs dramatically from the way it enters into general relativity, which is just one of the many challenges still facing physicists in their attempts to combine relativity and quantum mechanics into one unified theory of quantum gravity.

SYMMETRY

The universality of the laws of nature has fascinating mathematical origins and is linked with one of the most powerful ideas in science: symmetry. At a rudimentary level, everyone understands what is meant by a geometric shape that is symmetrical. A square is symmetrical because if you draw a line vertically down its centre, splitting it in half (or doing the same thing with a horizontal or diagonal line), then swapping the two halves around doesn't alter its shape. You also achieve the same result if you rotate it in multiples of 90 degrees. A circle has even more symmetry, because you can rotate it by any angle without changing its appearance.

In physics, symmetries can tell us something much deeper about reality than just the invariance of certain shapes when they are rotated or flipped. When physicists say that a physical system has a symmetry, they mean that some property of that system stays the same when something else changes. This turns out to be a very powerful concept. 'Global' symmetries are when laws of physics remain the same (there is no change in the way they describe some feature of the world) as long as some other change, or 'transformation', is applied equally everywhere. In 1915, Emmy Noether discovered that, wherever we see such a global symmetry in nature, we can be sure to find an associated law of conservation (a physical quantity remaining the same). For example, the fact that the laws of physics don't change when you move from one place to the next gives us the law of conservation of momentum, and the fact that the laws of physics don't change from one time to the next gives us the law of conservation of energy.

This has proven to be an extremely useful idea in theoretical physics and has deep philosophical

consequences. Physicists are always on the lookout for deeper, less obvious symmetries that are hidden in their mathematics. Noether's theorem tells us that we don't 'invent' the mathematics in order to have a way of describing the world, but rather, as Galileo observed, that nature speaks the language of mathematics, which is 'there', ready and waiting to be discovered.

The search for new symmetries has also helped physicists in their quest to unify the forces of nature. One such mathematical symmetry—that is not so easy to explain—is called supersymmetry. We do not yet know if this is a true property of nature, but if it is, then it could help us solve a number of mysteries, such as what dark matter is made of and whether string theory is the correct theory of quantum gravity. The problem is that this symmetry predicts the existence of a number of as yet undiscovered subatomic particles. Until we have experimental verification, supersymmetry remains just a neat mathematical idea.

Physicists have also learnt a lot—and picked up a stack of Nobel Prizes for their efforts—by

trying to find exceptions to the rules and laws that these symmetries give us, an idea known as 'symmetry breaking'. Have you ever sat at a circular dinner table in restaurant or at a fancy function and forgotten whether the side plate for your bread roll is to your left or your right? Before any of the guests at your table touches anything, the neatly laid out arrangement of plates, glasses and cutlery is symmetric. Etiquette aside, it doesn't really matter which side your bread plate is on, but as soon as someone makes a choice and (correctly) places a bread roll on the side plate to their left, the perfect symmetry is broken and everyone else can follow suit.

Symmetry breaking has helped physicists understand the building blocks of matter: the elementary particles, and the forces between them. The most famous example relates to one of the two types of force acting within the confines of atomic nuclei, known as the weak nuclear force. Until the 1950s, the laws of physics were thought to be exactly the same in a mirror reflection of our universe. This idea (swapping left and right)

is known as 'parity conservation' and is obeyed by the other three forces of nature: gravity, electromagnetism, and the strong nuclear force. But it turns out that the weak nuclear force, which is responsible for protons and neutrons transforming into each other, breaks this mirror symmetry. It does not lead to exactly the same physics when left and right are switched over. This violation of reflection symmetry now forms an important ingredient in the Standard Model of particle physics.

REDUCTIONISM

Much of modern science has been built on the idea that to understand some complex property of the world, we need to break it down to its basic parts, like taking a mechanical clock apart to see how all the gears and levers fit together to make it work. This view, that the whole is no more than the sum of its parts, is known as reductionism, and it has been a staple of many disciplines in science to this day. The idea goes back to the Greek philosopher, Democritus, and his notion of

atomism—that matter cannot be infinitely divided but is instead composed of basic building blocks. Later philosophers, like Plato and Aristotle, argued against atomism, believing that there had to be something missing, which they thought of as 'the form of the thing', and which had to be added to the substance itself. Take, for example, the form of a statue. Its meaning and its essence are more than just the stone it is made of. This vague metaphysical notion is not a part of modern physics. But thinking of things in this way helps to make a clearer argument against reductionism.

Let's take another example: water. We can study the properties of a molecule of H_2O as much as we want: the geometry of the bonds between the oxygen and hydrogen atoms and the quantum rules that govern this, the way water molecules stick together and arrange themselves, and so on. But we would not be able to *deduce* the property of 'wetness' of water by looking solely at its constituent parts down at the molecular level. This 'emergent' property only becomes apparent when trillions of water molecules come together in bulk.

Does this then imply that the whole is more than the sum of its parts, in the sense that there is some extra physics that we need to include to explain, for example, the bulk properties of matter? Not necessarily. The idea of emergence—that there are qualities of the physical world, like heat or pressure or the wetness of water, that do not have counterparts at the level of atomic physics—does not mean that there is more to a system than the sum of its parts, provided those emergent properties are still only built upon more fundamental concepts, such as the electromagnetic forces between subatomic particles in the case of water.

The reductionist enterprise continued when physicists in the nineteenth century attempted to understand the properties of complex systems that could not be explained by the simple laws of Newtonian mechanics. Towards the end of that century, James Clerk Maxwell and Ludwig Boltzmann developed two new subfields of physics—thermodynamics and statistical mechanics—which helped physicists to learn about systems made up of many pieces by look-

ing at them 'in bulk'. (We will look at these areas of physics more deeply in chapter 6.) Thus, while it is true that we cannot measure the temperature or pressure of a gas by looking at how its individual molecules vibrate and bump into each other, we still know that temperature and pressure are due to nothing more than the collective behaviour of individual molecules. What else can there be?

But while this simplistic reductionist line of thinking is not wrong—in the sense that there is no extra physical process that magically appears when we zoom out from the molecular scale—it is of limited usefulness when trying to describe the properties of a complex system. What we require is not 'new' physics, but 'more' physics, in order to learn about and understand how certain properties can emerge in a system from the collective behaviours of its constituents. The Nobel laureate Philip Anderson summed this view up in the title of a famous paper: 'More is different.'[8]

---

8 In this paper, published in 1972 [P. W. Anderson, Science 177 (4047): 393–96}], Anderson made his argument

But knowing that more physics is needed when we put the constituent parts (the particles, atoms and molecules) together to make up bulk matter is not the same as saying we know what that missing physics is. This becomes clear if we try to find a unified picture of the physical universe. We are still unable to derive the laws of thermodynamics from the Standard Model of particle physics, for example—or indeed to do the reverse, since it is not obvious which of these two pillars of physics is the more fundamental. And we are even further away from understanding more-complex structures, such as what dis-

---

against extreme reductionism. He used as an example the hierarchy of scientific disciplines arranged in a linear order, from physics, as the most 'fundamental' science, to chemistry to biology to psychology to the social sciences. This hierarchy did not imply, he claimed, that one subject is just an applied version of the one below, since 'at each stage, entirely new laws, concepts and generalizations are necessary, requiring inspiration and creativity to just as great a degree as in the previous one. Psychology is not applied biology nor is biology applied chemistry.' As an argument against reductionism, I regard this as somewhat weak. Whether or not a concept is fundamental does not depend on how profound it is, or how much inspiration or creativity was required to understand it.

tinguishes life from non-life. After all, you and I are still only made up of atoms, yet being alive is clearly more than just a matter of complexity, for a living organism is no more complex in terms of its atomic structure than an identical but recently deceased organism.

And yet . . . maybe we can dream of a time when we are able to have a single unified physical theory that underpins all natural phenomena. Until then, suffice it to say, a reductionist line of thinking only gets us so far, and we need to use different theories and models depending on what we are trying to describe.

THE LIMITS OF UNIVERSALITY

Despite our quest for laws of physics that are universal, the limits of reductionism point to the fact that sometimes the world can behave very differently at different scales and needs to be described and explained using the appropriate model or theory. For example, on the scale of planets, stars, and galaxies, gravity dominates everything—it controls the structure of

the cosmos. But it plays no role, that we can detect, down at the atomic scale where the other three forces (electromagnetism and the strong and weak nuclear forces) dominate. Indeed, probably the biggest unresolved problem in the whole of physics—one we will return to in chapter 5—is that the laws of physics that describe our everyday, so-called 'classical' world of matter, energy, space, and time simply don't work when we shrink down to the world of individual atoms, where the very different rules of quantum mechanics come into play.

Even at the quantum level, we often need to choose the appropriate model that is most applicable to the system we wish to study. We've known since the early 1930s, for example, that the atomic nucleus is made up of protons and neutrons; but in the late 1960s, it was discovered that these particles are not elementary, and are in fact made up of even tinier, more fundamental constituents: the quarks. This has not meant that nuclear physicists were forced to describe the properties of nuclei using quark models. A simplistic reductionist

approach might suggest that this is necessary for a deeper, more accurate description of the atomic nucleus. But that would not be very helpful. To a very good approximation, when describing the properties of nuclei, protons and neutrons behave as though they are structure-less entities and not composite systems of three quarks. So, while their properties and behaviour *must* ultimately be due to their deeper structure, this is not apparent or necessary if we wish to understand properties like the shape or stability of a nucleus. In fact, even within nuclear physics itself, a number of very different mathematical models are employed—each applying best to a certain class of nucleus; there is not a universal theory of nuclear structure.

This is what I mean by the world behaving differently at different scales of size, duration, and energy. While two of the wonderful things about physics are the universality of many of its theories and the way we can understand more about a system by digging deeper and under-standing how its parts relate to the whole, it is also true that we often have to choose the most

appropriate theory depending on the scale we are interested in. If you want to fix your washing machine, you do not need to understand the intricacies of the Standard Model of particle physics—even though washing machines, like everything else in the world, are ultimately made up of quarks and electrons. If we tried to apply our most fundamental theories about the quantum nature of reality to every aspect of our day-to-day lives, we wouldn't get very far.

Now that we have explored both the potential and the limitations of what physics can tell us—from the power of the mathematical symmetries underpinning our physical laws, to the sheer scale over which these laws can be applied, to the limitations of reductionism and universality—we are ready to get down to business. In the next chapter, I begin with the first of the three fundamental pillars of physics: Einstein's relativity.

SPACE AND TIME

In such a short book I am unable to cover all areas of physics, fascinating though so many of them are. Instead, I have distilled our current understanding of the physical universe down to three central pillars: three pictures of reality that come from very different directions. The first of these, introduced in this chapter and the next, is built on the work of Albert Einstein in the early twentieth century. It lays out our present understanding of the way matter and energy behave within space and time on the very largest scales due to the influence of gravity—an understanding that is encompassed in his famous general theory of relativity.

In order to paint Einstein's picture of the world, we must start with the canvas itself. Space and time are the substrates in which all events take place. However, such concepts are slippery.

Common sense tells us that space and time should be in place from the start—that space is *where* events happen and the laws of physics are acted out, while the inexorable passage of time is, well, just *is*. But, is our commonsense view of space and time right? An important lesson physicists must learn is to *not* always trust common sense. After all, common sense tells us that the Earth is flat, but even the Ancient Greeks understood that its sheer size meant we could not easily discern its curvature, but that there were simple experiments they could perform to prove that it was in fact a sphere. Similarly, everyday experience tells us that light has the properties of a wave and therefore cannot also behave as though it were made up of a stream of individual particles. If it were, how could we explain interference patterns? And yet it has been proven beyond doubt, through careful experiments, that our senses can deceive us when it comes to the nature of light. And when it comes to the quantum world, we must abandon many everyday notions based on simple intuition if we are to truly understand what is going on.

Learning not to always trust our senses is a valuable skill that physicists have inherited from the philosophers. As far back as 1641, René Descartes argued in his *Meditations on First Philosophy* that in order to know things about the material world that were absolutely true, he first needed to doubt everything, often despite what his senses were telling him. This doesn't mean that we cannot believe anything we are told or shown, but that, according to Descartes, those material things he judges to be true 'demand a mind wholly free of prejudices, and one which can be easily detached from the affairs of the senses'.[1]

In fact, long before even Descartes considered this, the medieval scholar Ibn al-Haytham began a philosophical movement in the early eleventh century known in Arabic as *al-Shukuk* (The Doubts), and he wrote extensively, particularly on the celestial mechanics of the Greeks, that

---

1 From the 1911 edition of *The Philosophical Works of Descartes* (Cambridge University Press), translated by Elizabeth S. Haldane, p.135.

one should question past knowledge and not take what one is told without evidence. This is why physics has always been an empirical science relying on the scientific method of testing hypotheses and theories through experimentation.

Nevertheless, some of the most important breakthroughs in physics have been the results of the logical conclusions drawn not from real experiments or observations, but from 'thought experiments', whereby the physicist considers some hypothesis and devises an imaginary experiment that can test its consequences. Such an experiment may or may not be possible to perform in practice, but it can still provide us with a valuable tool to learn about the world through the power of logic and reasoning alone. Some of the most famous thought experiments were conducted by Einstein and helped him develop his theories of relativity. Once his theories were fully developed of course they could be tested in real laboratory experiments.

When it comes to the meaning of space and time the difficulty we have is not surprising, for we are ourselves imprisoned within them, and it

is hard to free our minds from their confines and 'see' reality from the outside. And yet, incredibly, this is possible to do. In this chapter, I will outline our current understanding of the nature of space and time—a celebration of the debt we owe to Einstein and his two beautiful theories of relativity.

HOW DOES A PHYSICIST DEFINE SPACE AND TIME?

An important characteristic of Newtonian physics is that space and time have a real existence independent of the matter and energy that exist within them. But philosophers the world over had contemplated this idea long before Newton. For instance, Aristotle believed that empty space did not exist in its own right—that without matter there can be no space. Much later, Descartes argued that space was no more than the distance (or 'extension') between bodies. According to these two great thinkers, the space inside an empty box only exists because of the confines of the box—take away the walls of the

box, and the volume that was inside it no longer has any meaning.

But let us explore this example a little. What if you subsequently discovered that the box was sitting within the empty space of a larger one? Does the space within the smaller box, after its walls are removed, continue to exist now that it forms a part of the volume in the larger box? And must it therefore have been a real 'thing' all along? Imagine now that the empty smaller box—where, by empty, I mean truly devoid of anything: a vacuum—is *moving* in a vacuum contained within the larger box. Is the empty space inside the smaller box the same empty space as it moves, or is it occupying different parts of the space within the larger box? This is easy to answer if we replace the 'empty space' inside the sealed smaller box with water. As the box moves around within a larger volume of water, we can accept that it keeps the same water molecules inside it while displacing the water outside as it moves. But what if there is no water? And what if we now get rid of the physical walls of both boxes, and everything else in this imaginary universe, so

that all that is left is nothingness? Is that nothingness still something? Does this empty space exist ready to be filled with matter, or to be contained within the confines of a box? Maybe I am just asking the same question in different ways, but only because it is by no means a trivial one.

Isaac Newton believed that space has to exist in order for matter and energy to be contained within it and for events to take place within it. But space exists, he argued, only as an empty nothingness, independently of the laws of physics that govern the behaviour of matter and energy within it. For Newton, space is the canvas on which reality is painted. For without space—and time, of course—to fix events to, how would we be able to assign coordinates to locate events? Surely, they must happen at 'some point' in space and at 'some moment' in time. Without absolute space and time in place, how can we hope to anchor reality?

But was Newton right? The answer we can give today is both yes and no. (Sorry.) He was correct in the sense that space *is* real—it is more than just the gaps between things, as Descartes

had argued. But he was wrong about space having an absolute existence *independently* of what it contains.

These two statements sound contradictory . . . until you learn about Einsteinian relativity. Einstein proved that absolute space and absolute time do not exist as separate entities. But to appreciate why this notion is necessary, I need to introduce you to the first of his two theories of relativity.

EINSTEIN'S SPECIAL THEORY

Until Isaac Newton completed his work on the laws of motion, debates about the nature of time were considered to be the domain of philosophy and metaphysics rather than proper science. Newton described how objects move and behave under the influence of forces, and since all motion or change requires time to make any sense, time had to be included as a fundamental part of his mathematical description of the world. But Newtonian time is absolute and relentless; it flows at a constant rate, as though there were

an imaginary cosmic clock ticking off the seconds, hours, days, and years independently of the events and processes taking place in space. Then, in 1905, Einstein brought the Newtonian world crashing down by revealing how time is connected to space at a deep level.

Einstein's conclusion was that time is not absolute: it doesn't run at the same rate for everyone. If I see two simultaneous events—say, two flashes of light from sources on either side of me—then someone else moving past me at that very moment will not see them happening at the same time, but rather one slightly after the other. This is because the rate of flow of time for each of us depends on our state of motion relative to each other. This weird notion is one of the very first lessons of relativity theory and is called the relativity of simultaneity. Let us take a step back and look at these concepts more carefully.

Consider how sound waves travel to your ears. Sound is, after all, nothing more than the vibration of air molecules that pass on energy through their collisions. Without matter (air) there would be no sound. In space, no one can

hear you scream, as the byline to the 1980s movie *Alien* correctly pointed out.

Einstein's insight was to suggest that, unlike sound waves, light waves do not need a medium to carry them. His theory rested on two ideas (known as the principles of relativity). The first, originating with Galileo, states that all motion is relative and that there is no experiment that can be performed to show that someone or something is truly at rest. The second principle states that light waves travel at a speed that is independent of the speed of the source of the light. Both these ideas seem reasonable, until you dig a little deeper into their implications. Let us consider the second idea first—that light moves at the same speed for everyone—and carry out a simple thought experiment.

Imagine an approaching car on an empty country road. The sound waves from its engine will reach you ahead of the car since they are travelling faster, but their speed has to do with how quickly the vibrating air molecules can transmit them; they do not reach you any quicker if the car speeds up. What happens instead is that they

get compressed to shorter wavelengths. This is the well-known Doppler effect that we recognise as the change in pitch of the car as it finally reaches you and goes past. When it is receding, the waves of sound are emitted from a progressively longer distance away and therefore reach us stretched to longer wavelengths, and hence a lower pitch. So, while the *wavelength* of sound waves depends on the speed of their source, the *speed* of the waves themselves relative to us (how long they take to reach us) does not change unless *we* start moving through the air towards the approaching car. So far, so good, I hope.

Light is different. It does not need a medium to travel through, and with respect to which we can measure its speed. This means that no one has a privileged position in which they can say they are truly at rest and therefore can reliably measure the 'true' speed of light. Einstein concluded from this that we should all measure light to have the same speed regardless of how fast we are moving relative to each other. (Provided, that is, we are not undergoing any acceleration

or deceleration while measuring the speed of the light some distance away from us.[2])

Now consider two rockets approaching each other at constant speed close to that of light. But they have no reference point to argue over who is, or isn't, moving. An astronaut on board one rocket sends a pulse of light towards the oncoming second rocket, measuring the speed of the pulse as it travels away from him. Since he can quite legitimately claim to be at rest, floating in empty space, while the other rocket is doing all the moving, he should see the light moving away from him at its usual speed of just over one billion kilometres per hour,[3] and he does indeed see this. But at the same time, the astronaut in the second rocket can also legitimately claim to be floating stationary in space. She too

---

2   This is a technical detail. Basically, general relativity deals with non-inertial reference frames where spacetime appears curved due to gravity or acceleration. In such non-inertial frames, you only measure light to have a constant speed as it passes close to you.

3   The speed of light in empty space is 1.0792528488 billion km/hr.

will therefore expect to measure the speed of the light reaching her to be just over one billion kilometres per hour (since, like the sound waves from the car, the light's speed should not depend on the speed that its source is approaching her). And she does indeed measure light to have this speed. It would seem, therefore, that both astronauts measure the same light pulse to be travelling at the same speed, despite moving towards each other themselves at near light speed!

This strange nature of light turns out to be a property of the speed at which it can travel rather than of light itself—a speed that is the maximum possible in our universe and which stitches space and time together into one fabric. For the only way light can travel at the same speed for all observers regardless of how fast they themselves are moving relative to each other is if our concepts of distance and time change.

Here is another example. Imagine that you, on Earth, send out a series of light pulses, or flashes, into space to chase down a friend who has headed off in a very fast rocket—a powerful

futuristic one that can travel at 99 percent of the speed of light. You will measure the light pulses to be travelling away from you at one billion kilometres per hour and therefore slowly overtaking your friend's rocket at just 1 percent of the speed of light, in the same way that a car in the fast lane of a motorway that is travelling just little faster than one in the slow lane overtakes it at a speed that is the difference between their two speeds. But what does your friend in the rocket see if she tracks the overtaking light pulses? Relativity theory tells us that she will see them overtaking her at one billion kilometres per hour. Remember, the speed of light is constant, and all observers see it travelling at the same speed.

The only way for this to make sense is if time on board the rocket is ticking by at a slower pace than for you on Earth. That way, what you see as the slow overtaking of a light pulse past the rocket window, your friend sees as a light pulse flashing past, because very little time will have elapsed on the rocket's slower-ticking clock—although for your friend, the clock is ticking at a normal rate. Thus, one of the consequences of all observ-

ers seeing light moving at the same speed is that we all measure distances and times differently. And we do indeed see this: the constancy of the speed of light for all observers is a fact, verified experimentally over and over again, and without which our world wouldn't make sense.

The special theory of relativity resolves this counterintuitive situation beautifully by combining time and space in order to retrieve something that we can all agree on. Imagine the whole of space is contained within a vast rectangular three-dimensional box. To define an event taking place within the box we assign to it x, y, and z coordinates (indicating its position relative to the three axes of the box) along with a value for time (when the event took place). Common sense would tell us that the time value is quite different from the three numbers defining the event's location in space. But what if we could add a time axis to the three of space? It would need to be in a 'direction' that is at right angles to each of the three spatial axes, which is impossible for us to visualise. This would result in a combined four-dimensional volume of space-*plus*-time.

An obvious simplification to help us with this visualisation is to sacrifice one of the dimensions of space and collapse our 3-D volume onto a two-dimensional surface, leaving the freed-up third dimension to use as the time axis. Now, think of this static block of space and time as a giant loaf of sliced bread, where the time axis lies along the length of the loaf. Each slice of bread is a snapshot of the whole of space at a single moment, while successive slices correspond to successive times. This is known in physics as the block universe model. While it is only three-dimensional (two of space and one of time), we must not forget that it really represents a four-dimensional construct: 4-D spacetime. Mathematically, we don't have a problem dealing with four dimensions; it's just picturing them that is not possible.

Perceiving 4-D spacetime from the outside, we would experience the totality of existence, not only of all space, but of all times: past, present, and future, coexisting and frozen. It is an impossible viewpoint, an omniscient one, because in reality we are always trapped within the block

universe and we experience the passage of time as a steady crawl along the time axis, moving smoothly from one slice of the loaf to the next, like frames in a movie stacked alongside each other instead of end to end on a reel. The reason the concept of the block universe is so useful is because it allows us to understand our different perspectives according to relativity theory. Two observers moving at high speed relative to each other might each record two events—say, flashes of light—but they will not agree on how far apart those flashes are or the time interval between them. This is the price we must pay if we are to all see light moving at the same speed. Viewed within the four dimensions of the block universe, spatial distances and time intervals can be combined, so the separation between any two events, called the spacetime interval, will be the same for all observers. Their disagreement about distances and times, if treated separately, turns to be nothing more than different perspectives in spacetime. You and I can look at a cube from different angles so that what I see as its depth (the distance measured along my line of sight) will not appear the
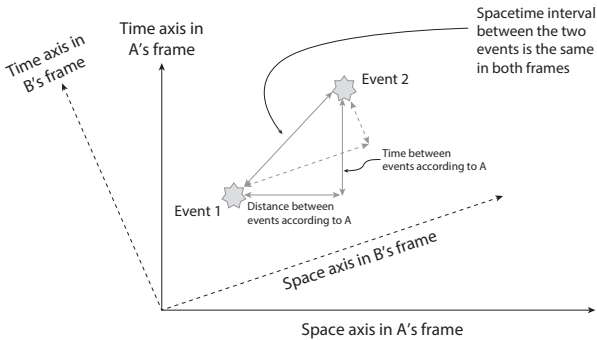
FIGURE 1. Events in spacetime—two observers, A and B, moving at high speed relative to each other, both see two events (flashes of light), which are separated in both space and time. They will not agree on the distance between the events, or on the time duration between them. This is because their space and time axes are different. But in 4-D spacetime (here two dimensions of space are ignored for simplicity) the (spacetime) interval between the two events in both frames is the same: the two right-angled triangles have the same hypotenuse even though each has a different space distance and time distance.

same to you if you see it face on. It depends on the angle at which we are looking at it. But we can nevertheless agree that it is a cube of equal-length sides, and that any differences are just down to our different perspectives. The same is happening in the 4-D block universe. We will always agree on spacetime intervals between events.

Einstein's relativity teaches us that we must view things within 4-D spacetime, in which both spatial and temporal distances become just a matter of perspective. No observer has the right to claim that their perspective of space and time is more correct than any other, because we will all agree once space and time are combined. Individual perspectives of space and time separately are relative, but combined spacetime is absolute.

EINSTEIN'S GENERAL THEORY

Just as the special theory merges space with time, Einstein's general theory of relativity links spacetime with matter and energy, which I will discuss further in the next chapter, to give a more profound explanation of the concept of gravity than that of Newton. According to Newton, gravity is an attractive force: an invisible rubber band between masses that pulls them together and acts instantaneously between them no matter how far apart they are. Einstein gives us a deeper and more accurate explanation: that the strength of

the gravitational pull that a body feels is a measure of the curvature of spacetime around it.

Again, this curvature is not something we can visualise. It is impossible to imagine flat 4-D spacetime, let alone when it's curved. For most everyday purposes, Newton's depiction of gravity as a force is a good enough approximation to reality, but its shortcomings become ever more evident when gravity gets a lot stronger, such as when we approach a black hole, or if we need to measure distances and times very accurately, such as onboard GPS satellites. In such cases, we are forced to abandon the Newtonian picture and fully embrace Einstein's vision of curved spacetime.

Since gravity is defined by the curvature of spacetime, this means that it influences the passage of time as well as the shape of space. For us, embedded within spacetime, this effect manifests itself as a slowing down of time, in a similar way to what we see when objects move close to the speed of light. The stronger the gravity, the slower a clock will tick compared to one far from the source of the field, in a 'flatter' region of spacetime.

Unfortunately for those who prefer complex ideas explained in plain language rather than dense mathematics, most attempts by physicists to describe how and why time runs more slowly in stronger gravity fall short of either explaining the phenomenon correctly or explaining it at all. But I will try.

Just as two people moving relative to each other will, according to special relativity, measure each other's clock ticking at a slower rate, a similar situation arises between the two observers if they are a fixed distance apart, but one of them is feeling a stronger gravitational pull—say, on the surface of the Earth, while the other is hovering far out in space. Again, the two of them will disagree on the time interval between events. As before, their clocks will tick at different rates: being deeper within the Earth's gravitational well, where there is more spacetime curvature, the Earth observer's clock will tick more slowly. However, unlike in special relativity, the situation here is no longer symmetrical, as she would see the clock out in space ticking more *quickly.* In a very real sense, gravity slows the flow of time.

We can say that the reason that a body 'falls' to Earth is because it always moves to where time runs the slowest—it is trying to age more slowly. Isn't that beautiful?

So much, then, for the effect of gravity on time. But what about space? What does general relativity tell us beyond the somewhat unhelpful remark that gravity 'causes space to curve'? Remember how both Aristotle and Descartes argued that, without any matter to fill it, space did not have an independent existence? Well, Einstein would take this a step further. According to his general theory, matter and energy *create* a gravitational field, and spacetime is nothing more than the 'structural quality' of this field. Without the 'stuff' contained within spacetime, there is no gravitational field and hence no space or time!

This may sound somewhat philosophical, and I suspect even some physicists will be uncomfortable with it. The problem is, in part, down to the way we teach physics. We tend to start with special relativity and 'flat' spacetime (because it is easier to teach and because

Einstein hit upon it first), then we progress on to the more difficult general relativity, in which this flat spacetime is filled with matter and energy, causing it to curve. In fact, conceptually we should think of it the other way around, beginning with matter and energy within spacetime. This way, special relativity is just an idealised approximation that only works when gravity is so weak that spacetime can be regarded as 'flat'.

The point I wish to make is a subtle one, and you can take solace from the fact that even Einstein himself did not fully appreciate its implications to begin with. Two years after completing the general theory, he wrote a popular science book (or 'booklet', as he referred to it) entitled *Relativity: The Special and the General Theory (A Popular Exposition)*, which was first published in German in 1916. Over the next four decades of his life, as he honed his understanding of what the maths told him about the universe, he would add appendices to this booklet. In 1954, the year before he died, he wrote his fifth and final appendix: two dozen

pages of prose containing some of the most profound ideas ever produced by the human mind.

To understand Einstein's thinking, we must understand the concept of a 'field' in physics. The simplest definition of a field is that it is a region of space containing some form of energy or influence, in which every point can be assigned a value that describes the nature of the field at that point. Think of the magnetic field surrounding a bar magnet. The field is strongest close to the poles of the magnet and becomes progressively weaker the further away in space from the magnet we get. The pattern of iron filings that arrange themselves along the magnetic field lines is simply their way of reacting to the field they are immersed in. But the point I wish to make sounds too obvious to warrant stating: *the magnetic field needs space to exist in*.

In stark contrast, the gravitational field, as described by Einstein and created by the mere existence of matter, is more than just a region of influence *within* space and time. It *is* spacetime.

Einstein went to great lengths in appendix 5 of his 'booklet' to clarify his thinking on this. In a new preface to the 1954 edition, he says:

> [S]pacetime is not necessarily something to which one can ascribe a separate existence, independently of the actual objects of physical reality. Physical objects are not in space, but these objects are spatially extended. In this way the concept of 'empty space' loses its meaning.

Then, in appendix 5, he clarifies this further: 'If we imagine the gravitational field . . . to be removed, there does not remain a space of the type (1) [i.e., flat spacetime], but absolutely nothing.' Flat spacetime, 'judged from the standpoint of the general theory of relativity, is not a space without field, but a special case . . . which in itself has no objective significance. . . . There is no such thing as an empty space, i.e., a space without field.' He concludes, 'Spacetime does not claim existence on its own, but only as a structural quality of the field.' Building on the ideas of Aristotle and Descartes, Einstein generalised the

notion that there is no space without material bodies and showed that there is no spacetime without a gravitational field.

Just like our magnetic field, the gravitational field is a real physical thing—it can bend, stretch, and undulate. But it is also more fundamental than the electromagnetic field: the electromagnetic field needs the gravitational field to exist, since without a gravitational field there is no spacetime.

EXPANSION OF SPACE

There is one final point I wish to make before I move on. A common confusion many have with the idea of spacetime curvature becomes apparent when physicists describe the expansion of the universe. If spacetime is one big static four-dimensional block, what does it mean when physicists talk about it expanding? How can something that includes time embedded within it expand? After all, the word 'expand' suggests something changing with time, but that something *contains* time! The answer is that the ex-

pansion of space that we observe through our telescopes does not involve any stretching of the time coordinate too. It isn't spacetime that is stretching, but rather only the three dimensions of space expanding as time moves forward. Although spacetime is in some sense democratic, with time as just one of the four dimensions, we can algebraically manipulate the equations of general relativity (by which I mean recast them in a slightly different form) so that all distances will now be multiplied by a 'scale factor' that increases as time moves forward and only space expands.

Remember also that this expansion only happens in the vast expanses *in between* the galaxies, because within the galaxies themselves the gravitational field that holds them together is strong enough to withstand the overall cosmic expansion. Galaxies are like the raisins embedded within a loaf of rising bread in the oven. The loaf expands, but the raisins themselves remain the same size—they just become more separated from each other.

In terms of the block universe, imagine that our local spacetime sits within a 'bread universe'

in which successive slices of the loaf, as we move along the time axis from past to future, get bigger. Floating outside of spacetime, you'd just see the static loaf with its increasing slice sizes. But from our vantage point trapped within the loaf (or within a figurative raisin within the loaf), all we can experience is successively larger slices, and so we see a point (a distant galaxy, say) moving further away from us as we move through the slices.

Despite all these deeply profound concepts, everything about spacetime that I have described in this chapter comes from just one of the three pillars of modern physics. But space, according to relativity theory, is smooth and continuous. If we zoom in, smaller and smaller, we will ultimately reach the domain of the second pillar of modern physics, quantum mechanics, where everything is fuzzy and subject to chance and uncertainty. What then happens to space and time at these tiniest of length scales and shortest of intervals? Will spacetime itself become grainy, like the pixels of an image magnified beyond its resolution? Maybe. We will come to that soon.

The block universe in relativity also says that we can think of time as static and unchanging, with past, present, and future coexisting as part of four-dimensional spacetime. But the third pillar of physics, thermodynamics, tells us that the idea of time as 'just another dimension' is inadequate. Thermodynamics describes the way systems change with time; more than that, it gives a directionality to time that is missing from the three dimensions in space. Independently of our own perception of time flowing in one direction only—born from the fact that we remember the past, live in the present, and anticipate the future—there exists an arrow of time that points from the past to the future, ruining the neat symmetry of the block universe.

But we are not yet ready to explore these next two pillars of physics. First we must fill our spacetime with stuff: matter and energy. The lesson from Einstein is that matter, energy, space, and time are all intimate companions. We will explore what this means in the next chapter.

CHAPTER 4

ENERGY AND MATTER

The general theory of relativity is mathematically encapsulated in what is known as Einstein's field equation (actually a set of equations that can be written together in a compact form on a single line). But equations always have two sides separated by an '=' sign, and the shape of spacetime is only half of the equation. I now wish to explore the other side.

Einstein's equation expresses how a gravitational field, or rather, the shape of spacetime, is determined by matter and energy. It is often said that his field equation shows how spacetime is curved by matter and energy and, at the same time, how matter and energy behave in curved spacetime. The point is that, just as matter and energy cannot exist without somewhere to exist

in, there would, equally, be no spacetime without matter and energy. So, let us explore what we know about the 'stuff' of the universe.

ENERGY

Energy is one of those concepts we all feel we understand intuitively. For example, we say we feel 'low on energy' if we're hungry, tired or unwell; conversely, if we're fit and well, we may feel 'energetic' enough to head for the gym. Sometimes, people use the term in a very unscientifi*c* way in phrases like 'I could feel the positive energy in the room,' or 'You are giving off a lot of negative energy'. In physics, the concept of energy indicates the *capacity to do work*; thus, the more energy something has, the more it is able to do, whether that 'doing' means moving matter from one place to another, heating it up, or just storing the energy for later use. The notion of energy has been used widely in physics for a couple of centuries, ever since it was found to be more useful as a concept than the admittedly more tangible notion of 'force'—since we can feel forces, but

we do not always have a direct sense of energy if it is not in the form of heat or light.

The definition of energy as the capacity to do work does nevertheless link it to the idea of a force, for when we use the term 'work' in physics, we generally mean the ability to move something against a resisting force. For example, I need energy to move a heavy piece of furniture along the ground against the force of friction, or to lift something above my head against the pull of gravity. Similarly, a battery expends energy pushing an electric current through a circuit against a conductor's resistance; and the heat energy stored in steam produces pressure to power turbines that transform this energy into electricity, which can then be used to produce mechanical work—or light and heat again.

Energy comes in many different flavours: a moving body has kinetic energy; a body in a gravitational field has stored potential energy; and a hot body has thermal energy due to the motion of its atoms. But while all of this is correct, it doesn't get to the heart of what energy actually is.

Let us begin with the *law of conservation of energy*, which states that the total amount of energy in the universe is constant. This follows, through Noether's theorem, from a deeper idea of time symmetry: that all the laws of physics are 'time translation invariant', which leads to the total energy of a physical process being conserved over time. This has led to profound new insights, such as the prediction of the existence of new elementary particles. The conservation of energy also tells us that perpetual motion machines are impossible, since energy cannot be continually conjured up from nowhere.

On the face of it, you might think that this is all there is to it: the total amount of energy in a system (indeed, in the entire universe) is conserved, even though it changes from one form to another. But there is something deeper about the nature of energy that I have not yet mentioned. In a rather loose sense, we can divide it into two types: *useful energy* and *waste energy*—a distinction that has profound consequences linked to the arrow of time. We know we need energy to run our world, to feed our transport and our

industries, to generate the electricity we use to light and heat our homes, to run our appliances and to power all our electronic devices. Indeed, energy is required just to sustain life itself.

Surely this cannot last forever. So, will we one day run out of useful energy? Zooming out, we can think of the entire universe as a wound-up mechanical clock that is slowly running down. But how can this be so if energy is always conserved? Why can't energy circulate indefinitely, changing from one form to another, but always there? The answer turns out to be down to simple statistics and probability, and what is known as the second law of thermodynamics. But if you don't mind, I will save that discussion until chapter 6. For the moment let us move on, from energy to matter.

MATTER AND MASS

Whenever we talk about the nature of matter, we also need to understand the concept of mass. At the most basic level, the mass of a body is a measure of the amount of 'stuff' it contains.

In everyday language, mass is often taken to mean the same thing as weight. This is fine on Earth since the two quantities are proportional to each other: if you double a body's mass you will also double its weight. But out in empty space, a body has no weight, even though its mass stays the same.

However, even mass does not always remain constant. The faster a body moves, the more its mass increases. This is not something you will be taught at school, and Isaac Newton would have found it astonishing, because it is yet another consequence of the nature of spacetime as elucidated by Einstein's Special Theory of Relativity. If you are wondering why we don't see this in everyday life, it is because we do not typically encounter things moving close to the speed of light, where the effect becomes noticeable. For example, a body moving at 87 percent of the speed of light, relative to some observer, will be measured by that observer to have double the mass it has when it is not moving, and a body moving at 99.5 percent of the speed of light will have ten times the mass it had when it was 'at

rest'. But even the fastest bullet only travels at 0.0004 percent of the speed of light, which means we generally do not experience relativistic effects or changes in moving bodies' masses.

The increase in the mass of a body as it reaches a significant fraction of the speed of light does not mean that it grows larger in size, or that the number of atoms it contains increases, but rather that it gains more momentum (making it harder to stop) than you might expect based simply on its 'at rest' mass. According to Newtonian mechanics, a body's momentum is the product of its mass and its speed, meaning that its momentum increases in proportion with its speed—you double its speed and its momentum doubles. But Newtonian mechanics says nothing about masses increasing when a body is moving. Special relativity gives us a different (and more correct) 'relativistic' formula for momentum, which is no longer proportional to a body's speed. In fact, momentum becomes infinite when a body reaches light speed.

This is a useful way of understanding why nothing can travel faster than light (another of the predictions of Special Relativity). Think of

the energy needed to make a body move faster. At low speeds, this energy gets transferred into kinetic energy (energy of motion) as the body speeds up. But the closer the body gets to the speed of light, the harder it gets to make it go even faster, and the more of the energy being put into it gets used to increase its mass instead. This notion leads to the most famous equation in physics: $E = mc^2$, which links mass (m) and energy (E) together (along with the square of the speed of light, c) and suggests that the two quantities are transformable into each other. In a sense, mass can be thought of as frozen energy. And because the speed of light squared is such a large number, a small amount of mass can be converted into a large amount of energy, or conversely, a large amount of energy freezes into very little mass.

Therefore, we see that the law of conservation of energy is more accurately generalised to the law of conservation of energy *and* mass: the total amount of energy plus mass in the universe is constant over time. Nowhere is this notion clearer or more important than in the subatomic world, where $E = mc^2$ led to an understanding of

nuclear fission and the unlocking of the energy of the atomic nucleus. And it is $E = mc^2$ that lies behind half a century of accelerator laboratories in which beams of subatomic particles are smashed together at ever higher energies to create new matter—new particles—out of the energy of the collision. But there are rules associated with what sort of matter particles can be created from energy, and we will discuss some of them in the next section.

THE BUILDING BLOCKS OF MATTER

From the moment, over a century ago, when Ernest Rutherford, with the help of Hans Geiger and Ernest Marsden, probed the interior of atoms for the first time, by aiming alpha particles at a thin gold leaf and watching how many passed through it and how many bounced back, physicists have been obsessed with delving ever deeper into the subatomic world. They first revealed the structure of atoms themselves—electron clouds surrounding a tiny, dense nucleus. Then, they looked inside the nucleus itself to discover that it is made

of smaller building blocks, protons and neutrons. And eventually, they zoomed in even deeper to reveal the elementary quarks hidden within the protons and neutrons. To give you an idea of scale, if an atom were blown up to the size of a house, then the volume within a proton or neutron in which the quarks are confined would be the size of a single grain of salt. And remember that atoms themselves are incredibly tiny: you can fit more atoms into a single glass of water than there are glasses of water in all the oceans of the world.

At school, we learn about the electromagnetic force in the form of electrical or magnetic attraction or repulsion, but it plays an even more crucial role down at the atomic scale. Atoms bond together in all sorts of combinations, to make simple molecules and complex compounds and ultimately the huge variety of different materials we see around us. But how the atoms bind together comes down to the way their electrons arrange themselves around the nuclei, which is of course the very essence of chemistry, and this binding together of atoms to make up the stuff of our world is almost entirely due to the electromagnetic force

between the electrons. In fact, together with gravity, the electromagnetic force is responsible, either directly or indirectly, for nearly all the phenomena we experience in nature. On the microscopic scale, materials are held together by the electromagnetic forces between atoms. On the cosmic scale, it is gravity that holds matter together.

Within the atomic nucleus is a very different world. Since nuclei are made up of two types of particles, positively charged protons and neutral neutrons (collectively called nucleons), electromagnetic repulsion between the protons should force nuclei apart; and gravity is far too weak at this tiny scale to be of any use. And yet the constituents of nuclei are held together tightly. This is thanks to a different force that works as a glue to stick protons to neutrons and even protons to protons, despite the repulsion of their positive charges. It is called the strong nuclear force and is felt most strongly between the even tinier constituents of protons and neutrons: the quarks themselves, which are bound together by 'force carrier particles' called gluons. Thus while quarks are attracted to each other by exchanging

gluons, a quark and an electron interact via the electromagnetic force (because they both have electric charge) by exchanging photons.

The quantum rules governing the structure, shapes and sizes of atomic nuclei are very complicated and will not be discussed here. Ultimately, however, it is the interplay of the *repulsive* electromagnetic force between the positively charged protons and the *attractive* nuclear force between all nucleons (which is itself a remnant of the 'strong' force—the internal 'gluonic' attraction between the quarks within the nucleons) that contributes to the stability of nuclei, and hence of atoms, and hence of all the matter around us, including us.

There is also another force—the fourth and final (known) force of nature, which is also—mostly—confined within atomic nuclei. It is known simply as the weak nuclear force, and it arises from the exchange of W and Z bosons between certain particles (in the same way that quarks exchange gluons and electrons exchange photons). Like the strong nuclear force, this weak force also acts over very short ranges, and we

do not to see its effects directly. However, we are very familiar with the physical processes triggered by this force, as it causes protons and neutrons to transform into each other, which in turn leads to beta radioactivity: charged particles ejected from nuclei. Beta particles come in two types: electrons and their antimatter partners, positrons, which are the same as electrons but with opposite charge. The process is quite simple: if a nucleus has an imbalance between the number of protons and neutrons it contains, leading it to become unstable, then one or more protons or neutrons will transform into the other to redress the balance. In the process, either an electron or a positron is created and ejected (ensuring that electric charge is conserved). Thus, a nucleus with too many neutrons will undergo beta decay in which a neutron will change into a proton, and an electron is emitted, its negative charge cancelling out the positive charge of the created proton as is required (since the original neutron had no charge). Conversely, an excess of protons prompts one of them to convert into a neutron plus a positron that carries off the

proton's positive electric charge, leaving a more stable nucleus behind.

Protons and neutrons each contain three quarks, which come in two types (or 'flavours') known somewhat unimaginatively as 'up' and 'down'. These two flavours carry different fractions of electric charge. A proton contains two up quarks, each with a positive charge equivalent to two-thirds of the negative charge of an electron, and one down quark with negative charge that is just one-third of the electron's. Added together they make up +1, the correct positive charge of the proton. The neutron, on the other hand, consists of two down quarks and one up quark, so its total charge is zero.

In total, six different flavours of quarks exist, each with a different mass. As well as the up and down quarks that make up atomic nuclei, the other four are called 'strange', 'charm', 'top', and 'bottom'—all arbitrarily chosen names. These quarks are heavier than the 'up' and the 'down,' but only exist fleetingly. Finally, in addition to *electric* charge, quarks also have another

property known as *colour* charge, which relates to the strong nuclear force and helps explain the way quarks interact with each other.[1]

Electrons belong to another class of particles, called leptons, of which there are also six types. Along with the electron, there is the muon and the tau (short-lived heavy cousins of the electron) and three types of neutrinos (very light, almost undetectable particles that are formed during beta decay). Leptons do not feel the strong nuclear force and do not carry colour charge.

To summarise, according to our current understanding, the Standard Model of particle physics tells us that there are, overall, two kinds of particles: the matter particles (the fermions), which include six flavours of quarks and six leptons; and the force carrier particles (the bosons),

---

1 As well as the threesomes of quarks making up nucleons, they can also come in pairs (strictly, quarks and antiquarks) to make up another class of particle called mesons. We still do not know for sure if quarks can combine to make more exotic composite particles, such as so-called tetraquarks, which would be made up of two quarks and two antiquarks, or pentaquarks, with four quarks and one antiquark.

which include the photon, the gluons, the W and Z, and of course the Higgs, which I will discuss later on.

If all this sounds unnecessarily complicated, then you'll be relieved to hear that for most practical purposes it needn't be so. Everything you see: all the stuff that makes up our world, including our own bodies, and everything we see out in space: the Sun, the Moon, and the stars, is all made of atoms, and all atoms are, in turn, made up of just two kinds of particles: quarks and leptons. Indeed, all atomic matter consists of just the first two quark flavours (the up and down), plus one of the leptons (the electron). Although you may be surprised to know that the most common matter particle is the neutrino.

A BRIEF HISTORY OF MATTER AND ENERGY

So how and when did all this matter come to be in the first place? To understand this we need to zoom back out again and explore the cosmos on the largest scales.

That our universe is expanding has been known for almost a century. Astronomers observed the light from distant galaxies to be stretched towards the red end of the electromagnetic spectrum (or, redshifted), indicating that these galaxies are moving away from us. In fact, the further away galaxies are, the greater the redshift of their light and so the faster they must be moving. However, seeing galaxies receding in every direction does not mean that we occupy a privileged position in the centre of the universe. Rather, it means that *all* galaxies are moving away from each other, because the space between them is stretching. Note that this expansion does not apply *within* clusters of galaxies, like our Local Group: the Milky Way, Andromeda, and a handful of smaller galaxies, which are close enough to each other to be gravitationally bound and thus able to resist the expansion of space.

But what, you may ask, has the expansion of the universe to do with the origin of matter and energy? Well, this expansion is one of the most compelling pieces of evidence we have for the Big Bang—the moment 13.82 billion years ago

when our part of the universe was born in a state of incredibly high temperature and density. Put simply, if the universe we see is expanding now, with the galaxies flying apart, then everything must have been closer together in the past. At some point in time, if we go back far enough, all the matter, as well as the space containing it, must have been squeezed together. There is therefore no location in the universe that we could travel to, plant a flag, and claim that the Big Bang happened there. The Big Bang happened everywhere in the universe. And just to confuse you further, if the universe is infinite in size now (as it may well be), then it would have to have been already infinite in size at the Big Bang (since you cannot expand something finite to make it infinite—unless you have infinite time to do it!). That the Big Bang happened *everywhere* within already infinite space, rather than at some particular 'place', is an important concept to grasp.

A more up-to-date, and conceptually more logical, take on this notion is that the Big Bang we refer to is only a 'local' event. It created just the visible universe we are able to discern, whereas

the entirety of the infinite universe contains other distant regions of space beyond what we could ever see, and which had their own big bangs. This is one of the ways of explaining the idea of a multiverse, which I will come to in chapter 8.

There is plenty of other evidence to support the Big Bang theory, too, such as the relative abundances of the light elements. About three-quarters of the mass of all the matter we see in the universe is in the form of hydrogen, and one quarter is helium (the next lightest element).[2] Only a tiny amount exists of all the other elements, most of which were made in stars long after the Big Bang. This dominance of hydrogen and helium in the universe is predicted by the Big Bang theory and is exactly what we observe. And the great thing is that we don't need to travel around the universe to determine this composition. The light we collect in our telescopes car-

---

2 Note that I used the word 'mass' here. In terms of numbers of atoms in the universe, about 92% are hydrogen and only 8% are helium (because helium is four times the mass of hydrogen).

ries within it the telltale signature of the distant atoms that have produced it or that it has passed through on its journey to Earth. The fact that we can learn about the ingredients of the universe just by studying the light that reaches us from space is one of the most beautiful notions in science.

The other piece of evidence in support of the Big Bang—the discovery of which in 1964 finally confirmed the theory beyond reasonable doubt—is the existence of the so-called cosmic microwave background (or CMB) radiation. This ancient light that fills all of space originated at a time, not long after the Big Bang, when neutral atoms first formed, during a period in the universe's history called the 'era of recombination'. It took place 378,000 years after the Big Bang, when space had expanded and cooled enough for positively charged protons and alpha particles[3] to capture electrons and form

---

3 Alpha particles are nuclei of helium, the next lightest element after hydrogen. They consist of four nucleons: two protons and two neutrons.

hydrogen and helium atoms. Before this, electrons would have been too energetic to stick to the protons and alphas to make neutral atoms; consequently, photons (the particles of light) couldn't travel very freely without bumping into and interacting with these charged particles, so the whole of space would have taken on a foggy glow. But, once the universe cooled enough for atoms to form, space became transparent and the photons were set free. This light has been travelling across the universe in all directions ever since.

This first light has also been losing energy as space expands, but not by slowing down, since light always travels at a constant speed. Instead, it is the wavelength of the light that has been stretched with the expansion of the space it is moving through, so that today, billions of years later, it is no longer in the visible part of the spectrum, but in the form of microwaves. Astronomers have measured this microwave radiation and found it to correspond to a temperature of deep space of a little less than three degrees above absolute zero, a value that agrees with the

prediction of the Big Bang theory—which, by the way, was made before the measurement.

But let us go back to an even earlier time in our universe's life, long before atoms were even formed. It began as a stupendously hot bubble of energy and, within a trillionth of a second, had cooled enough for subatomic particles, quarks and gluons, to form—condensing out from this energy as space expanded. To begin with, these particles were very energetic and roamed around unconfined in a hot soup called the quark-gluon plasma, at a temperature of trillions of degrees Celsius. Then, when the universe was a mere millionth of a second old, they began to clump together to form protons and neutrons (along with other heavier particles). Matter then went through various stages of evolution in those first few seconds, with different particles forming and disappearing. It is here that we encounter one of the biggest outstanding unanswered questions in physics: the mystery of the missing antimatter.

A few years after Paul Dirac predicted its existence in 1928, antimatter was discovered by Carl Anderson in cosmic rays: high-energy particles

from space that collide mainly with oxygen and nitrogen molecules in the Earth's upper atmosphere to produce a shower of secondary particles, including the electron's antiparticle, the positron. We now know that all elementary matter particles (the fermions) have mirror-image antimatter partners.[4] When an electron and positron come into contact, they completely annihilate each other with their masses combining and transforming into pure energy via $E = mc^2$.

The reverse of this annihilation process is also continuously taking place down at the tiniest scales. If we could magnify the quantum realm, we would see particles and their antiparticles popping in and out of existence all the time in a constant exchange between matter and energy. Thus, a photon, which is no more than a lump of electromagnetic energy, can transform itself into an electron and a positron in a process known as pair creation. But, in the very early, dense universe, when particles and antiparticles were

---

4 The other type of particle, the force carriers like photons, are called bosons and technically do not have antiparticles.

appearing and disappearing, matter for some reason came to dominate over antimatter. The fact that we are here at all indicates that this must have been so. We have yet to understand what happened to the 'missing antimatter' that, luckily for us, gave rise to the extra profusion of matter we see today.

A few minutes after the Big Bang, the conditions were right for protons (nuclei of hydrogen) to fuse together to make helium[5] plus a tiny amount of element number three, lithium. But as the universe expanded further, the temperature and pressure dropped below the threshold for heavier nuclei to be formed via the fusion of lighter ones. This is because, for nuclear fusion to take place, the fusing nuclei must be energetic enough to overcome the mutual repulsion of their positive charges, but below a certain matter density and temperature, this no longer happens.

A little later, after the era of recombination, atoms began to clump together under the

---

5 Technically, there are several steps here, including the beta decay of protons into neutrons.

influence of gravity—and I am holding back on the vital role played by dark matter here, but will say more about it in chapter 8—and primordial gas clouds (proto-galaxies) began to form, and denser clumps of gas within them were squeezed together even more dramatically by gravity until they heated up sufficiently for the process of fusion to begin once again. Stars ignited, and the thermonuclear reactions taking place inside them produced new elements: carbon, oxygen, nitrogen, along with many of the other elements we find on Earth.

Most of this first generation of stars in the universe no longer exist, since they would have exploded as supernovae long ago, spewing much of their elemental contents out into space, leaving behind compacted matter in the form of neutron stars or black holes. Heavier elements—that is, anything beyond iron in the periodic table—are only created during violent events such as novae, supernovae, and neutron star mergers. The hotter and the more extreme the conditions in a star, the further the nucleosynthesis process is able to go, and the heavier the elements that can be

formed, such as silver, gold, lead, and uranium. This is because the interiors of stars only reach the required temperature and density to make heavier elements during these final intense moments of their lives, when they are densely compressed, while at the same time violently shedding their outer layers.

The matter ejected from exploding stars blends with interstellar gas, which can clump together again to form a new generation of stars. The fact that we find such heavy elements on Earth tells us that our Sun is such a second-generation star (at least). It is why you might have heard it said that we are all quite literally made of stardust, for indeed it is within stars that many of the atoms in our bodies were made.

Now that I hope I have given you a sense of how matter was formed in the universe and the intimate relationship between matter and energy, space and time, we are ready to plunge down into the microcosm, a world of the very small that cannot be described by the general theory of relativity. It's time to explore the second pillar of physics: quantum mechanics.

CHAPTER 5

THE QUANTUM WORLD

In 1799, Joseph Banks, president of the august Royal Society of London, founded a new establishment: the Royal Institution of Great Britain, with the aim of introducing 'useful mechanical inventions and improvements' and 'teaching [the general public] courses of philosophical lectures and experiments'. Ever since then, the Ri, as it is commonly known today, has continued to put on public lectures and events, including its Friday Evening Discourses—public lectures delivered in its Faraday lecture theatre—which have been an integral part of its programme since they were set up by Michael Faraday himself in 1826. I have had the honour of giving two of these, the last one in 2013, when I talked about the subject of this chapter: quantum mechanics.

Quantum mechanics is seen, quite rightly, as the most fascinating, yet at the same time most mind-boggling and frustrating scientific theory ever devised by humankind. In a particular segment of my lecture at the Ri, I discuss the famous 'two-slit experiment', which describes what the American physicist Richard Feynman called the 'central mystery of quantum mechanics'. After outlining just how astonishing the results of the two-slit experiment are—subatomic particles, fired one by one through a screen with two narrow slits in it, behaving as though they each travel through *both* slits at once, and giving rise to an interference pattern on a second screen— I issued a challenge to my audience. If anyone were able to come up with a 'commonsense' account of how this is possible, they should get in touch with me, as they will no doubt be up for a Nobel Prize.

I said this as a lighthearted joke—safe in the knowledge that no one has ever found a simple explanation of this classic result despite many decades of debate and hundreds of ingenious tests, leading physicists to reluctantly conclude that whatever is going on really does *not* have

a commonsense explanation. This really is the way matter behaves in the quantum world, and we just have to accept it. I also assumed, when I cast down the proverbial gauntlet, that I was addressing just the few hundred members of the Ri audience that Friday evening. But the Ri posts much of its educational material online, which included my lecture; and since then I have received hundreds of emails from amateur scientists claiming to have solved this central quantum mystery and suggesting that maybe physicists have forgotten to consider this or that mechanism or detail.

I used to respond, but I confess that I don't any more. So, let me make amends for my lack of correspondence with those folks who continue to puzzle over the mysteries of quantum mechanics, and describe some of its most important, and non-intuitive, features. In this chapter we will take a brief look at what this second pillar of modern physics tells us about the microcosm. Having devoted my own research career, now careering into its fourth decade, to its study and application, first in nuclear physics and more

recently in molecular biology, you won't be surprised to hear that I regard quantum mechanics as the most powerful and important theory in all of science. After all, it is the foundation on which much of physics and chemistry is built, and it has revolutionised our understanding of how the world is built from the tiniest of building blocks.

A QUANTUM MECHANICS PRIMER

The status of physics towards the end of the nineteenth century appeared to be complete. It had produced Newtonian mechanics, electromagnetism, and thermodynamics (which I will talk about in chapter 6) and showed that together these three areas of physics successfully described the motion and behaviour of all everyday-size objects and pretty much all phenomena we encounter around us, from cannonballs to clocks, storms to steam trains, magnets to motors, and pendulums to planets. Collectively, the study of all these things is referred to as Classical Physics, and it is still predominantly what we are taught in school.

However, classical physics, while still pretty good, is not the whole story. When physicists turned their attention to the microscopic constituents of matter—atoms and molecules—they discovered new phenomena they couldn't explain with the physics they knew. It seemed that the laws and equations they were using no longer applied. Physics was about to undergo a seismic paradigm shift.

The first major theoretical breakthrough—the concept of the 'quantum'—was made by the German physicist Max Planck. In a lecture in December 1900, he proposed the revolutionary idea that the heat energy radiated by a warm body is linked to the frequency at which its atoms vibrated, and consequently that this radiated heat is 'lumpy' rather than continuous, emitted as discrete packets of energy, which became known as quanta. Within a few years, Einstein had proposed that it wasn't just Planck's radiation that was emitted in lumps, but that all electromagnetic radiation, including light, came in discrete quanta. We now refer to a single quantum of light—a particle of light energy—as a photon.

Einstein's proposal, that light is quantum in nature, was more than just a hunch. It explained one of the biggest outstanding scientific mysteries of the time, called the photoelectric effect—a phenomenon in which light, when shone on a metal surface, can knock electrons out of the metal's atoms. This effect could not be explained if light were a wave because if so, then increasing the intensity (the brightness) of the light would mean increasing its energy, and we would expect the electrons knocked out from the metal to fly off faster. But, they don't. There are just more of them. But if the energy of the light is proportional to its frequency rather than its intensity, as Einstein proposed, then increasing its frequency (for example, from visible to ultraviolet) would cause the electrons to be knocked out with more energy. And, conversely, keeping the frequency (colour) of the light the same but increasing its brightness would just mean that *more* photons would be produced, and more electrons knocked out. This is exactly what is seen in experiments, and Einstein's explanation fitted beautifully.

And yet there was, and still is, plenty of contrary evidence suggesting that light is made up of waves rather than a stream of particles. So which is it? Is light a wave or a particle? The answer, frustratingly, flying in the face of intuition and common sense, is that it can behave like either, depending on how we look at it and the sort of experiment we devise to probe it.

And it is not just light that has this schizophrenic nature. Particles of matter, such as electrons, can exhibit a wavelike nature, too. This general notion, tested and confirmed for almost a century now, is known as wave-particle duality and is one of the central ideas of quantum mechanics. This does not mean that an electron is both a particle and a wave at the same time—but rather that, if we set up an experiment to test the particle-like nature of electrons, we find that they do indeed behave like particles. But if we then set up another experiment to test if electrons have wavelike properties (such as diffraction or refraction or wave interference), we see them behaving like waves. It's just that we cannot carry out an experiment that would

show both the wave and particle nature of electrons *at the same time.* It is absolutely vital to stress here that, while quantum mechanics correctly predicts the outcomes of such experiments, what it does *not* tell us is *what an electron is*—only what we see when we carry out certain experiments to probe it. The only reason this no longer drives physicists crazy with exasperation is that we have learnt to accept it. This balance between how much we can simultaneously know about an electron's particle nature (its position in space) and its wave nature (how fast it is travelling) is governed by Heisenberg's uncertainty principle, which is regarded as one of the most important ideas in the whole of science and a foundation stone of quantum mechanics.

The uncertainty principle puts a limit on what we can measure and observe, but many people, even physicists, are prone to misunderstanding what this means. Despite what you will find in physics textbooks, the formalism of quantum mechanics does not state that an electron cannot *have* a definite position and a definite speed

at the same time, only that we cannot *know* both quantities at the same time. A related common misunderstanding is that humans must play some kind of crucial role in quantum mechanics: that our consciousness can influence the quantum world, or even bring it into existence when we measure it. This is nonsense. Our universe, all the way down to its elementary building blocks at the quantum scale, existed long before life began on Earth—it wasn't sitting in some fuzzy limbo state waiting for us to come along, measure it, and make it real.

By the mid-1920s, physicists were beginning to realise that the concept of quantisation is more general than just the 'lumpiness' of light or the 'waviness' of matter. Many physical properties, familiar to us as continuous, are, in fact, discrete (digital rather than analog) once you zoom down to the subatomic scale. For example, the electrons bound within atoms are 'quantised' in the sense that they can only have certain specific energies and never energies in between these discrete values. Without this property, electrons would continuously leak energy while orbiting

the nucleus,[1] meaning that atoms would not be stable and complex matter, including life, could not exist. According to nineteenth-century (pre-quantum) electromagnetic theory, negatively charged electrons should spiral inwards towards the atom's positively charged nucleus. But their quantised energy states prevents this from happening. Certain quantum rules also define which energy states the electrons occupy and how they arrange themselves within atoms. As such, the rules of quantum mechanics dictate how atoms can bind together to make molecules, making quantum mechanics the foundation of the whole of chemistry.

Electrons are able to jump between energy states by emitting or absorbing the correct amount of energy. They can drop to a lower state by emitting a quantum of electromagnetic energy (a photon) of exactly the same value as the difference in energies between the two states

---

1 The term 'orbit' here is in fact wrong, since atoms are not miniature solar systems and electrons are not localised particles like tiny planets going round the sun.

involved. Likewise, they can jump to a higher state by absorbing a photon of the appropriate energy.

The sub-microscopic world, down at the scale of atoms and smaller, therefore behaves very differently from our familiar everyday world. When we describe the dynamics of something like a pendulum or tennis ball, or a bicycle or a planet, we are dealing with systems comprising many trillions of atoms, which are far removed from the fuzziness of the quantum realm. This allows us to study the way these objects behave using classical mechanics and Newton's equations of motion, the solutions of which are an object's precise location, energy or state of motion, all knowable simultaneously at any given moment in time.

But if we wish to study matter on the quantum scale, we must forgo the mechanics of Newton and use the very different mathematics of quantum mechanics. Typically, we would solve Schrödinger's equation to calculate a quantity called the wave function, which describes not the way an individual particle move along definite path, but the way its 'quantum state' evolves in time. The wave function can describe

the state of a single particle or group of particles and has a value that provides us with the *probability* of, say, finding an electron with any given set of properties or location in space *if we were to measure that property*.

The fact that the wave function has value at more than one point in space is often wrongly taken to mean that the electron itself is physically smeared out across space when we are not measuring it. But quantum mechanics does not tell us what the electron is doing when we are not looking—only what we should expect to see when we *do* look. If you are not reassured by this statement, you are not alone. It is not meant to reassure you (or to discourage you, for that matter); it is simply stating what all physicists agree on when it comes to the meaning of quantum mechanics.

Beyond this, there is a whole host of different ways of explaining the nature of the quantum world. These are known as the 'interpretations' of quantum mechanics, and the arguments between advocates of these different views have raged for as long as quantum mechanics has been around and show no sign of abating.

WHAT DOES IT ALL MEAN?

Despite its tremendous success, if we dig a little deeper into what quantum mechanics tells us about the microcosm, we could easily lose our minds. We ask ourselves, 'But how can it be so? What am I not 'getting'?' The truth is, no one really knows for sure. We do not even know if there is any more to 'get'. Physicists have tended to use terms like 'strange', 'weird', or 'counter-intuitive' to describe the quantum world. For, despite the theory being powerfully accurate and mathematically logical, its numbers, symbols and predictive power are a façade hiding a reality we find difficult to reconcile with our mundane, commonsense view of the everyday world.

There is, however, a way out of this predicament. Since quantum mechanics describes the subatomic world so remarkably well, and since it is built on such a complete and powerful mathematical framework, it turns out we can manage just fine by learning how to use its rules in order to make predictions about the world and

to harness it to develop technologies that rely on those rules, leaving the hand-wringing and head-shaking to the philosophers. After all, this laptop I am typing on would not exist were it not for the development of quantum mechanics that allowed us to create modern electronics. But if we take this pragmatic attitude, we must accept that we become no more than quantum *mechanics* ourselves—practitioners and technicians who do not care *how* or *why* the quantum world behaves the way it does, but who simply accept it, and move on. Every fibre of my being tells me that this should not be enough for a physicist. Is it not the job of physics to *describe* the world? Quantum mechanics, without an interpretation of its equations and symbols, is just a mathematical framework that allows us to calculate and predict the results of experiments. That should not be enough. Physics should be about *explaining what our results tell us about how the world really is*.

The fact that many physicists will not agree with this statement is a failing that can be traced back to one of the greatest thinkers in the history

of science: the father of quantum mechanics, Niels Bohr. So influential was he that even as I write this I feel a sense of guilt—that I am betraying one of my great heroes. And yet I must stand true to my convictions. Without doubt, Bohr's philosophical views have shaped the way generations of physicists think about quantum mechanics, but they have also, in the eyes of an increasing number, discouraged and stifled progress. Bohr argued that it is wrong to think that the task of physics is to find out how nature *is*—or to know the 'real essence of the phenomena'—but rather to concern itself only with what we can *say* about nature: the 'aspects of our experience'. These two opposing views, the first ontological and the second epistemological, can in fact *both* be correct: what a physicist should be able to say about nature, even at the quantum scale, should be the same as how nature is, or as close to it as we can get, but always trying to edge closer. This 'realist' view is one that I have always found myself siding with in the end, despite having serious doubts now and again.

On the other end of the scale, there is a danger lurking in the shadows should we overstress the weirdness of quantum mechanics rather than focussing on its power and success as a scientific theory. For doing so attracts the attention of charlatans just as a bright light attracts moths. The undeniably inexplicable predictions of quantum mechanics—such as entanglement, i.e., separated particles being instantaneously linked across space—have over the years provided fertile ground for all manner of pseudo-scientific nonsense, from telepathy to homeopathy. Generations of physicists have been trained to follow Bohr's pragmatic dogma—known as the Copenhagen interpretation of quantum mechanics, named after the city of Bohr's famous Institute for Theoretical Physics where so much of the earlier mathematical foundations of the theory were laid out in the mid-1920s—in part to avoid the sort of philosophical musings that can spill over into new-age baloney.

Like all physics students for generations, I was taught quantum mechanics by first being

introduced to its historical origins and the work of Planck, Einstein, Bohr, and others. But my education quickly moved on to the mathematical techniques (the toolkit) I needed to make use of the theory. And along with the maths, I learnt a pile of concepts named after the theory's founding fathers: Born's rule, Schrödinger's equation, Heisenberg's uncertainty principle, Pauli's exclusion principle, Dirac notation, Feynman diagrams . . . the list goes on. But while all of this is important to know if we are to make sense of the quantum world, what I was not taught were the arguments and philosophical debates that took place between all these great physicists, which lasted throughout their lives and which, to a large extent, remain unresolved.

Much of the interpretational difficulty with quantum mechanics revolves around the so-called 'measurement problem': How does the ephemeral quantum world come into sharp focus when we carry out a measurement? Where is the boundary between the quantum and the classical worlds—between things that do not have well-defined properties when left

to their own devices, and the reassuring solidity of what we measure and see? Many of the founding fathers—men such as Niels Bohr, Werner Heisenberg, and Wolfgang Pauli—believed it was pointless to worry about such matters and advocated following the Copenhagen philosophy I described earlier. They were happy to split the world in two, quantum behaviour and classical behaviour, without tying down how one transitioned into the other upon measurement. To them, quantum mechanics worked and that was enough. But this positivist attitude can hinder the advance of science. While it may well help lead us to a better understanding of some phenomena and even to developing new technologies, it does not help us to truly *understand*.[2]

The history of science is littered with examples of this sort of attitude. One of the most

2 Of course, my Copenhagenist colleagues would vehemently disagree with me here. They would argue that they do understand all there is to be understood about what quantum mechanics can and cannot tell us and that it is the realists who refuse to accept or understand this.

obvious comes from ancient cosmology. For two millennia, from antiquity until the birth of modern science, there was an almost universal hegemony and acceptance of the geocentric model of the universe: that the Earth is at the centre of the cosmos and that the Sun, along with all the planets and stars, orbits around us. A positivist back then would have argued that since this model works so well in predicting the motion of heavenly bodies, it was unnecessary to look for alternative explanations for how or why they moved across the sky in the way we see. Indeed, there was a time when the geocentric model was more accurate at matching astronomical observations than the correct, and much simpler, Copernican heliocentric model. But interpreting a theory in a particular way 'just because it works' is intellectual laziness, and certainly not in the true spirit of what physics should be about. The same should be true of quantum mechanics. The renowned quantum physicist John Bell once famously said that the aim of physics is to *understand* the world, and 'to restrict quantum mechanics to be exclusively about pid-

dling laboratory operations is to betray the great enterprise'.

Sadly, too many physicists, even today, do not grasp this—yet another argument for why philosophy is not just pointless navel gazing, but can contribute to the advancement of science. If you were to conduct a poll among quantum physicists (at least those who care about such matters), you would find that a significant, though decreasing, fraction will still adopt the pragmatic Copenhagen view. But a growing number see it as an abdication of the role of physics and instead subscribe to one of a number of alternative interpretations—a list that includes such exotic-sounding ideas as the many worlds interpretation, the hidden variables interpretation, the dynamical collapse interpretation, the consistent histories interpretation and the relational interpretation—and I have left out a number of others. No one knows which, if any, of these different ways of describing reality at the quantum scale is the correct one. They all work; they all make, so far, the same predictions of the results of experiments and

observations,[3] and all emerge from the same mathematics. Sometimes, advocates of these different interpretations can defend them dogmatically, treating their favourite version almost like religions, which is not how science is going to progress.

And yet, slow progress *is* being made in trying to understand the quantum world. Experimental techniques are becoming ever more subtle, and certain explanations are being ruled out. The hope is that one day we will indeed find out how Nature really does perform her quantum trickery. If this sounds sensible to you, there are plenty of physicists who will disagree. The positivists argue that science is nothing more than a tool for predicting the outcome of experiments and that those who worry about what quantum mechanics tells us about reality by reading more than they should into its mathematics are indeed better suited to doing philosophy instead. In fair-

---

3 Although some realist interpretations, such as spontaneous collapse models, do make predictions that others do not, and so are in principle testable.

ness, not all advocates of this positivist, Copenhagenist view of reality have been dismissive of attempts to dig deeper. In the early 2000s a new anti-realist interpretation called Quantum Bayesianism (or Qbism) appeared, whose proponents see reality as being entirely subjective and down to personal experience. Critics have even likened it to solipsism.

Choosing an interpretation of quantum mechanics should be more than just a matter of philosophical taste. The fact that they all make the same predictions about the world does not mean that all these interpretations are equivalent or that we are free to choose the one we like the best on a whim. Explaining some aspect of reality through physics is a two-step process. First, we find the mathematical theory, which of course may or may not be correct. But if we believe it is right—like Einstein's field equations of general relativity or Schrödinger's equation in quantum mechanics—then we next need ways of interpreting, or explaining, what the mathematics means. These are the stories we attach to the maths. Without them, we cannot

connect our symbols and equations, however aesthetically pleasing we find them, to the physical universe we observe. And we need to find the right story just as much as the right mathematical theory.

The different interpretations of quantum mechanics paint very different pictures of reality: either there are parallel universes (the many worlds interpretation) or there are not; either there is a physical nonlocal quantum field (the pilot wave hidden variables interpretation) or there isn't. Nature does not care about our petty squabbles regarding the correct interpretation of quantum mechanics—it gets on with doing things the way it does, and exists independently of our perceptions. If we have a problem with agreeing on how the quantum world behaves, then that is our problem. Einstein believed this. He was a realist, too. He believed that physics should be about describing how the world really is, and if there is more than one description that fits the mathematics of quantum mechanics, then we should not be satisfied. I feel I am in good company in this regard.

ENTANGLEMENT, MEASUREMENT, AND DECOHERENCE

That said, even Einstein could get things wrong on occasion. One of the most profound and inexplicable predictions of quantum mechanics is the idea of entanglement. In the quantum world, two or more particles can be linked across space instantaneously in a way that almost defies logic. Technically, this is known as nonlocality, and it can be encapsulated in the idea that what happens 'over here' can instantaneously affect, and be affected by, what happens 'over there'. We say that the two particles are described by the same 'quantum state': the same wave function. Einstein always felt uncomfortable about nonlocality and entanglement, deriding it as 'spooky action at a distance', and refused to accept that any communication between subatomic particles could travel across space faster than light, as that would violate special relativity. But in principle, particles located at opposite ends of the universe can still be connected in this way. Entanglement was shown by the quantum pioneers to follow

naturally from their equations, and experiments carried out in the 1970s and '80s confirmed that Einstein was wrong on this: we now know empirically that quantum particles really can have an instantaneous long-range connection. Our universe really is nonlocal.

Today, many researchers working in fields such as quantum optics, quantum information theory, and even quantum gravity see a profound link between entanglement and the central problem of measurement in quantum mechanics. We must first acknowledge that a quantum system—say, an atom—is in reality part of its surrounding world, and so treating it as isolated is not strictly correct. Instead, we must include in our calculations the influence of its surrounding environment. Such an 'open quantum system' presents us with a much more complex problem to solve, but at the same time, it allows us to make some headway in understanding what it means to carry out a measurement on a quantum system beyond what Niels Bohr referred to simply as 'an irreversible act of amplification' as the way to describe how quantum fuzziness is

crystallized into reality when we carry out an observation.

In fact, it is now clear that the environment surrounding a quantum system, such as an atom, can itself do the 'measuring'. We don't require a conscious observer. We can think of the atom as becoming ever more entangled with its surroundings, such that its quantum nature leaks out into the environment like heat dissipating from a warm body. This leaking out of the ephemeral quantum behaviour is known as decoherence and is part of an active area of study at the moment. The stronger the coupling between the quantum system and its environment, the more entangled it becomes, and the faster its quantum behaviour disappears.

Whether or not this process fully explains the measurement problem is still a matter of debate in some quarters. The thorny issue of how to solve the measurement problem and the boundary between the quantum world and the large classical world was first made famous by Erwin Schrödinger in the mid-1930s, when he came up with his famous thought experiment.

Despite being one of the pioneers and founders of the field, Schrödinger tried to highlight his own misgivings about the meaning of quantum mechanics. He asked what would happen if we were to shut a cat in a box with a radioactive substance and a container with lethal poison. All the while the box is closed, we cannot say whether or not a particle has been emitted by the radioactive material, triggering a mechanism that releases the poison, killing the cat. All we can do is ascribe probabilities to the two likely outcomes: when we open the box, either a particle has been emitted and the cat is dead, or it hasn't, and the cat is still alive. But according to the rules of quantum mechanics, and as long as the box is closed, the subatomic particle obeys the laws of the quantum world, and we must regard it as being in a quantum superposition of having been both emitted and not emitted at the same time.

But now, within the closed box, the fate of the cat rests on this quantum event. Schrödinger argued that since the cat is itself made of atoms, albeit trillions of them, each a quantum entity,

it too should exist in quantum superposition: a state of being both dead and alive simultaneously. However, we only ever see one definite outcome when we open the box to look. That is, the cat is either dead or alive, never in this limbo state.

A sensible way of resolving the issue is to assume that such quantum superpositions decohere away into their surroundings and therefore do not survive for long when we consider complex macroscopic objects like cats, which are never in two states at once, even before we open the box to check. In fact, while an isolated radioactive atom must be described as being in a superposition of having both decayed and not decayed until observed, it is surrounded by a complex environment of air, Geiger counter, and cat, all of which it rapidly becomes entangled with, so that the both-at-once option doesn't survive.

So, has the problem been solved? And do the two options of dead *or* alive cat now reflect nothing more than our own ignorance of its fate, until we open the box? If not, then we are still left with the mystery of what physical

process is taking place when we open the box. What has happened to the option we don't see? Subscribers to the many worlds interpretation of quantum mechanics believe there is a neat and simple explanation for this. They argue that there are now two parallel realities in which each option is realised. What we find when we open the box reflects which reality we exist in.

Other physicists, not prepared to accept the idea of a potentially infinite number of parallel realities, have come up with a range of alternative interpretations that still demand the existence of an objective reality in the absence of measurement, but all of which contain some strange aspect of reality hidden somewhere. For example, another way of interpreting quantum theory was first developed by the French physicist Louis de Broglie in the 1920s and then improved several decades later by David Bohm. According to this interpretation, the quantum world is made up of particles guided by waves. Their properties are hidden from us—and are called hidden variables—but

describe a quantum world without any of the fuzziness of the standard Copenhagen picture of reality. Rather than an electron itself exhibiting both wavelike or particle-like properties depending on how we measure it, there are both waves *and* particles, but it is only the particles that we ever detect. A small but dedicated community of physicists around the world feel that this so-called de Broglie–Bohm theory has much to offer, but it remains a largely unexplored option among the corpus of quantum interpretations.

Fascinating though it is, I will leave this discussion here, since many other books cover it in greater depth than I have space for. In any case, I leave the issue of the interpretation of quantum mechanics unresolved, since that is where we stand at the moment.

Having focussed thus far on the basic building blocks of matter and energy, the spacetime in which they exist, and the quantum nature of reality underpinning it all, I have ignored some equally fundamental concepts in physics that emerge when large numbers of particles come

together to make up complex systems. So, let us leave the world of the very small behind for now and zoom out again to investigate what happens with the emergence of complexity, and explore such profound ideas as order, chaos, entropy, and the arrow of time.

CHAPTER 6

THERMODYNAMICS AND THE ARROW OF TIME

As we leave behind the quantum world, with its randomness and uncertainty, our familiar Newtonian world comes back into sharp focus. The steaming, swirling cup of coffee on the table, the ball that just bounced into the back garden from next door, or the jet flying high overhead are all, if you think about it, made of matter and energy assembled into systems of varying levels of complexity. So, if we want to understand the physics of the world we see around us, we need to understand how particles interact and behave in large collections. The area of physics that helps us to understand the behaviours of large numbers of interacting bodies is statistical mechanics.

You might also recall that, while familiarizing ourselves with matter and energy in chapter 4, we touched on the fact that energy can transform from one form into another while the total amount of energy in a system remains the same. The energy of a bouncing ball flips constantly between its potential energy, due to its height above the ground, and its kinetic energy of motion. At the very top of its bounce, it is entirely in the form of potential energy, and just before it hits the ground, when the ball is moving at its fastest, this potential energy will have been transformed to kinetic energy. This all sounds fairly straightforward—but we also know that a ball will not keep bouncing forever: it loses energy in the form of heat, from friction with the air and from collision with the ground. This change from kinetic energy to heat is fundamentally different from the transformation between kinetic and potential energy because it is a one-way process. We would be astonished if we saw the ball suddenly regain its bounce, without any external help.

So why should this be so? Where does this 'one-way-ness' come from?

The answer is that a ball loses its bounce for the same reason that heat always flows from a warm cup of coffee to the colder surrounding air and never back again, and why the sugar and cream in the coffee never un-dissolve and un-mix. Welcome to the field of thermodynamics—the third major pillar of physics (along with general relativity and quantum mechanics). While statistical mechanics describes how a large number of particles interact and behave in a system, thermodynamics describes the heat and energy of the system and the way these change over time. As you will see, these areas of study are highly interconnected, and so physicists often learn about them together. We will look at them together, too.

STATISTICAL MECHANICS AND THERMODYNAMICS

Consider a box full of air in which all the molecules are bouncing around randomly. Some are moving quickly, while others are slower. But if the box is maintained at a fixed temperature

and pressure, then the total amount of energy it contains remains constant. This energy is distributed among its molecules in a very particular way: the total available energy is spread out according to a simple statistical rule. Suppose you inject some hotter air (faster-moving molecules) into the box and then leave it alone: the random collisions of these new molecules with the original cooler ones will cause their energy to be distributed. The hot molecules will slow down while simultaneously causing others to speed up. Eventually, the air will settle back down into a new equilibrium. This time, the most likely energy of any molecule will be slightly higher than it was before, and the overall temperature in the box will have been raised a little.

The way the energy in the box is spread among the molecules is called a Maxwell-Boltzmann distribution—after two of the greatest scientists of the nineteenth century, who developed the field of statistical mechanics. 'Distribution' refers to the shape of the curve on a graph linking the varying speeds of the molecules to the num-

ber having each speed. Or, put another way, it is the line that links points corresponding to the probability than any molecule will have a given speed. There will be a particular speed that it is most likely for molecules to have, corresponding to the highest point on the curve, with speeds faster or slower being less likely; and the shape of the distribution changes as the temperature of the box increases, with the peak in the probability distribution moving towards higher speeds. When the molecules have settled down to a Maxwell-Boltzmann distribution, we say that the air in the box has reached thermodynamic equilibrium.

The tendency towards a statistical state of equilibrium is associated with a very important concept in physics, known as entropy. The entropy of a system, if left alone, will always increase: that is, a system will always relax from a 'special' (ordered) state to a less special (mixed-up) one. Physical systems unwind, cool down, and wear out. This is referred to as the second law of thermodynamics, and at its heart it is no more than a statement of statistical inevitability: if left

Maxwell-Boltzmann distributions
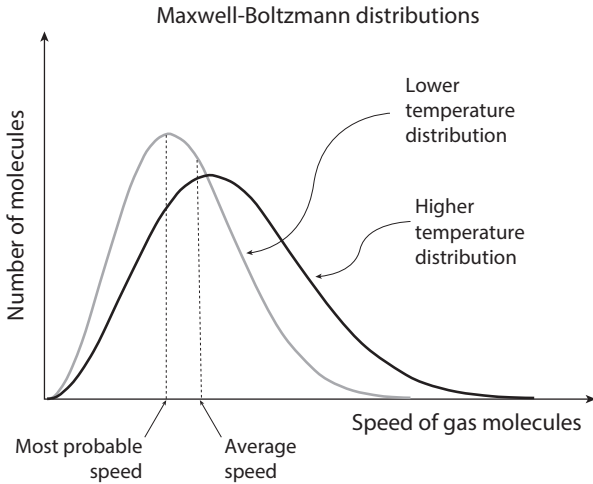
FIGURE 2. Maxwell-Boltzmann distribution—molecules of gas in a box will distribute themselves evenly and share their energy until they reach thermal equilibrium. The curve of number of molecules versus their speed is known as a Maxwell-Boltzmann distribution and has a peak at the most probable speed. This peak moves to higher speeds as the overall temperature of the gas increases. Note that the most probable speed is not the same as the average speed, since there are more particles with speeds greater than the peak value.

alone, everything always eventually returns to a state of equilibrium.

Imagine that all the molecules of air in our box start off clustered in one corner. The entropy of the box in this initial state is low, since its con-

tents are in a special, more ordered state. If left alone, the random motion of these molecules will cause them to spread out quickly to fill the whole box until their distribution reaches equilibrium. Just as the speeds of the hot molecules eventually settle into a state of thermodynamic equilibrium, the air in the box goes from a state of low entropy to a state of high entropy as it spreads out. When the air molecules are evenly distributed throughout the box, entropy will be at a maximum.

Here is an even simpler example. An ordered pack of cards in which each suit is separated and arranged in ascending order is said to have low entropy. It is in a highly ordered state, which is ruined if we shuffle the pack—we say that its entropy increases. By shuffling further, it is overwhelmingly more likely that the cards will get even more mixed up than it is for them to return to their original ordered arrangement. This is because the unshuffled pack is a very unique arrangement of the cards, whereas there are very many ways for the cards to be mixed up. So it is much more likely that shuffling will go in one direction—from unshuffled to shuffled, from low entropy to high entropy.

A more interesting definition of entropy is as a measure of something's ability to expend energy in order to carry out a task. When a system reaches equilibrium, it becomes useless. A fully charged battery has low entropy, which increases as the battery is used. A discharged battery is in equilibrium and has high entropy. This is where the distinction between useful energy and waste energy comes in. When a system is ordered and in a special (low-entropy) state, it can be used to carry out useful work—like a charged battery, a wound-up clock, sunlight, or the chemical bonds between carbon atoms in a lump of coal. But when the system reaches equilibrium, its entropy is maximised, and the energy it contains is useless. So, in a sense, it is not energy that is needed to make the world go around, it is low entropy. If everything were in a state of equilibrium, nothing would happen. We need a system to be in a state of low entropy, far from equilibrium, to force energy to change from one form to another—in other words, to do work.

We consume energy just by being alive, but we can see now that it has to be of the useful,

low-entropy kind. Life is an example of a system that can maintain itself in a state of low entropy, away from thermal equilibrium. At its heart, a living cell is a complex system that feeds (via thousands of biochemical processes) on useful, low-entropy energy locked up in the molecular structure of the food we consume. This chemical energy is used to keep the processes of life going. Ultimately, life on earth is only possible because it 'feeds' off the Sun's low-entropy energy.

The second law of thermodynamics and the relentless march of entropy applies to the entire universe, too. Imagine that our box of air is now a cloud of cold gas expanded to the size of a galaxy. If a group of molecules in the gas randomly drift closer together than average, then the mutual, very weak gravitational attraction between them might be enough to pull them closer together so that they form a denser clump of gas than average.[1] The more gas molecules clump

---

1 Of course, if we are dealing with small numbers of molecules, then gravity is never going to play a role in controlling their behaviour. Only when vast numbers of them are involved can their cumulative mass have a gravitational influence.

together, the more effective gravity becomes at attracting more molecules. This gravitational clumping process was the reason stars formed: vast clouds of gas collapsed together until these regions were dense enough for thermonuclear fusion (of hydrogen into helium) to begin, and stars ignited. This can appear puzzling when you first think about it, because the process of clumping together seems like it is resulting in a tidier, more ordered, and more 'special' state, and hence the end state should have lower entropy than when all molecules are spread out evenly. So, has gravity caused the entropy of the gas to decrease and the second law of thermodynamics to be violated?

The answer is no. Whenever matter clumps together gravitationally its entropy increases, for the same reason that the entropy of a ball increases as it rolls down a hill due to the pull of the Earth's gravity. Think of this clumping like a stretched spring being released, or a clock unwinding, their entropy increasing as they lose the ability to do useful work. Thus when the molecules of gas in a certain part of the cloud

Molecules of gas that are unevenly spread out will increase in entropy by spreading out evenly within the box to reach thermal equilibrium.

But if bodies are massive enough for gravity to play a role then they can also increase their entropy by doing the opposite and gravitationally clumping together.
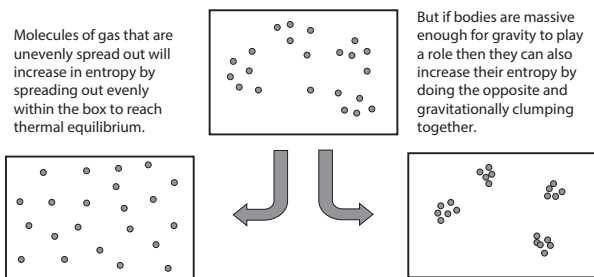
FIGURE 3. Increasing entropy—particles in a box that are slightly out of equilibrium (low entropy) can increase their entropy either by redistributing themselves back to equilibrium or clumping together under gravity. Either way, they increase their entropy and satisfy the second law of thermodynamics.

find themselves, by chance, temporarily closer together than when evenly spread out, this represents a temporary departure from maximum entropy. For entropy to increase again, to satisfy the second law, these molecules can do one of two things. They can either drift apart again, back to their original state of thermal equilibrium, or they can go the other way and clump together due to their mutual gravitational attraction. Either way, their entropy increases.

You should now be asking: What would cause such a drift away from maximum entropy to start

with? Wouldn't that drift itself violate the second law? The answer is that the matter and energy in our universe did not start in a state of thermal equilibrium, but rather in a very special, low-entropy state set by the conditions at the Big Bang itself. These initial conditions, down at the quantum level, seeded spacetime with irregularities, which became writ large on the fabric of the cosmos as the universe expanded, so that a certain amount of lumpiness was automatically built into the distribution of matter. As the universe continued to 'unwind', matter that was close enough together to feel the pull of gravity eventually clumped together to form stars and galaxies. The molecules of hydrogen and helium gas in space fell together into the gravitational wells of the stars, causing an increase in entropy as they did so. But, crucially, this entropy did not reach a maximum—stars are not systems in thermal equilibrium, but remain reservoirs of low entropy, with the thermonuclear fusion reactions within them releasing excess energy in the form of light and heat. It is this low-entropy energy from our own star, the Sun, that makes

life on Earth possible. Plants make use of it during photosynthesis to create biomass, locking useful low-entropy energy in the molecular bonds of their organic compounds, which can then be accessed by other living creatures, including eventually humans, that consume the plants as food.

The Earth itself also has a store of useful energy that, together with the Sun's energy, drives its climate, while the gravitational energy of the Moon and Sun control the tides of the oceans—all of which can provide us with other useful reservoirs of low entropy that we can tap into. For example, water at the top of a waterfall drops under the pull of gravity so that its potential energy is converted into kinetic energy, which we can make use of to drive hydroelectric power plants generating electricity for our use. There will of course always be some loss in efficiency—the second law needs to see some overall increase in entropy in the form of waste heat.

But there is something far more profound going on here than simply the transformation of energy from one form to another.

A DIRECTION TO TIME

If a physical system—including the entire universe—must always move from an ordered state of low entropy to a disordered state of high entropy, then this gives us a direction to the flow of time itself: the second law of thermodynamics allows us to distinguish between past and future. This might sound a little strange; after all, you don't need the second law to tell you that yesterday was in the past. You have a memory of the events of that day stored in your brain, though the events themselves are gone forever. Whereas tomorrow is unknown to you—it has yet to happen. This arrow of time pointing from past to future is, we feel, an intuitively more fundamental property of reality on which the second law of thermodynamics sits. In fact, it is the other way around: think of the second law of thermodynamics as the *origin* of time's arrow. Without the second law there would be no future or past.

Imagine watching a movie of our box of air (and let us imagine that the molecules of air are big enough for us to see). They will be bouncing

around, colliding with each other and with the walls of the box, some of them moving faster and others slower. But if the air is in thermal equilibrium, then we would not be able to tell whether the movie is being run forwards or backwards. Down at the scale of molecular collisions, we cannot see any directionality to time. Without an increase in entropy and a drive to equilibrium, all physical processes in the universe could happen equally well in reverse. However, as we saw, this tendency of the universe and everything in it to unwind towards thermal equilibrium is entirely down to the statistical probability of events at the molecular level progressing from something less likely to happen to something more likely, according to the laws of thermodynamics. The directionality of time pointing from past to future is not mysterious; it's just a matter of statistical inevitability.

With that in mind, even the fact that I know the past but not the future is no longer so strange. As I perceive the world around me, I increase the amount of information stored in my brain, a process which, as my brain is doing work, produces

waste heat and increases my body's entropy. Even our very ability to distinguish between the past and the future is, from a thermodynamical perspective, no more than our brains obeying the second law.

DETERMINISM AND RANDOMNESS

The above ideas may leave you feeling uneasy—and justifiably so. Surely, the difference between the past and future is more than just the statistical drive of randomly colliding molecules toward equilibrium, or the difference between an unshuffled and shuffled pack of cards. After all, the past is fixed—we remember only one course of events: one history. In contrast, the future is open to us with its infinite possibilities.[2] Most of the events that will happen tomorrow will be unexpected, and my day could unfold in myriad dif-

---

2 Of course, some things are more likely than others. . . . I am almost entirely certain that the Sun will rise tomorrow, and that I will be a day older; and I am pretty sure I won't wake up with the sudden ability to speak fluent Japanese or to run 100 metres in under ten seconds.

ferent ways, depending on the coming together of millions of different factors. So, is there in fact a difference between past and future, at a level deeper than simple statistics, which is based on the notion that we have one past but many possible futures? Put another way, is our fate sealed or is our future governed by chance? Is the future fixed or yet to be determined? These are age-old philosophical questions, touching on the nature of free will itself.

When physicists talk about a process as being 'deterministic', they are usually referring to the concept of 'causal' determinism: the idea that past events cause future events. But if this is the case, then nothing is left to chance, and everything that happens does so for a reason, because of what happened just before it: cause and effect. In principle, therefore, the state of the entire universe at the present moment can be traced back in time step by step, all the way to the Big Bang. And if this is true, then surely events in the present fix events in the future so that, *in principle*, we should be able to predict that future. And the term 'events' here also includes the firing of

neurons in our brains that define our thought processes and hence our decision making. After all, our brains are also made of atoms. There is no extra magical ingredient that exempts them from the laws of physics.

In a universe in which everything is predetermined, we would have no free choice with regard to our actions and decisions, since there is only one version of the future, just as there was one version of the past. (Remember I discussed Einstein's block universe idea in chapter 3.) But the order of events, the past causing the future and not the other way around, is driven by the second law of thermodynamics, without which the events we have labelled 'future' ones could just as likely have caused the 'past' ones.

But if this is the case, how is it that we are unable to predict the future with any degree of confidence? After all, even our most powerful supercomputers cannot tell us for sure if it is going to rain next week. In the case of the weather, the reason is straightforward. If you think about the sheer complexity of what we are trying to model and the number of variables

we would need to know very precisely—from temperature variations in the atmosphere and the oceans to air pressure, wind direction and speed, solar activity, and so on—in order to make an accurate prediction, you will see that the task becomes increasingly difficult the further into the future we want to make a forecast. So, while meteorologists can confidently predict whether it will be sunny or overcast tomorrow, predicting if it is going to rain on this date next year is impossible. Crucially, this does not mean that such knowledge couldn't *in principle* be known— since in a deterministic universe the future is already preordained—it is just that, *in practice*, we would need to know the current conditions of the Earth's climate to astonishing accuracy and have stupendous computational power to feed in all the data to make a precise simulation that could then be evolved mathematically to give a reliable prediction.

It is this chaotic *unpredictability* that give rise to the famous 'butterfly effect': the idea that the tiny, seemingly inconsequential disturbance of the air caused by the flapping of a butterfly's

wings on one side of the world could gradually develop and grow until it dramatically affected the course of a hurricane on the other side of the world. This does not mean that there is a specific butterfly to which we can trace the cause of a hurricane, but rather that *any* tiny changes to the initial conditions can give rise to widely varying outcomes if we continue to evolve the system in time.

The equations of physics describe a deterministically evolving world. Knowing the precise initial conditions of a system (where each constituent particle is and how it is moving at a given moment in time, and fully understanding the forces between all the particles) would allow us to compute how that system evolves in a *perfectly deterministic way*. Cause and effect. The future could (in principle) be laid bare for us.

The problem, of course, is that we can never do this in practice. This inability to know or control the initial conditions of a system, as well as all other continuing influences, to infinite accuracy can be seen even in systems much simpler

than the weather. The toss of a coin cannot be exactly repeated in order to achieve the same outcome time and again. If I toss a coin and get 'heads', it's too difficult for me to repeat the toss and make it spin the same number of times so as to definitely land 'heads' again. In a deterministic universe such as ours, our destiny is entirely mapped out, yet we are unable to predict it with any confidence.

But what about quantum mechanics? Isn't that where true randomness and indeterminism enter at a fundamental level? Does quantum mechanics not rescue us from the bleak determinism of a preordained, fixed future in which we feel we are no longer making free choices but are just cogs in an orderly clockwork universe? The truth is, we have no clear answer to this question yet. We must also take care to distinguish unpredictability from indeterminism. It is quite true that the probabilistic nature of the quantum world means that events are *unpredictable*: that we cannot know in advance exactly where an electron will be, or in which direction it is spinning, or precisely when a radioactive atom will

decay. All we can do with quantum mechanics is assign probabilities to the outcomes of different measurements. However, while this unpredictability *might* be down to true indeterminism, the mathematics of quantum theory does not require this. Indeterminism is an interpretation we impose on the mathematics to describe what we measure. For example, most cosmologists favour the many worlds interpretation of quantum mechanics in which everything is fully deterministic.

There is another way in which unpredictability and the appearance of randomness come into physics, and that is through the phenomenon of chaotic behaviour. Chaos appears in nature when there is an instability within a system, such that tiny changes to the way the system evolves over time can quickly grow. There's that butterfly effect again. Sometimes even simple systems following simple, deterministic physical laws can behave in highly unpredictable and complex ways that seem to be truly random. But unlike in the quantum domain, where we don't know whether unpredictability is due to true indeter-

minism or not,[3] the unpredictability of a chaotic system is not—despite initial appearances—due to true randomness.

There is also a fascinating flip side to chaos theory: that simple rules, applied repeatedly, can lead to seemingly random behaviour, but then sometimes go on to produce beautiful structures and complex patterns of behaviour that look highly ordered. Unexpected complexity emerges where there was none before, while never violating the second law of thermodynamics. The field of science dealing with this sort of emergent behaviour is known as complex systems, and it is beginning to play a major role in many exciting areas of research, such as biology, economics, and artificial intelligence.

In summary, then, it may well be that our universe is indeed completely deterministic, and any unpredictability about its future evolution is entirely due to the shortcomings in our own ability to know with certainty what will happen next.

---

3 Since it depends on which interpretation of quantum mechanics we choose.

This could either be because, at the quantum level, we cannot observe the state of a system without disturbing it and altering the outcome, or because we cannot, in practice, ever have *complete* knowledge of a system, and the build-up of uncertainties means we can never be sure what the future holds.

WHAT THEN IS TIME?

Having taken a brief look at determinism and randomness in physics, let us return once more to the central theme of this chapter, namely the direction of time emerging from thermodynamics. Note that we have now been confronted with three *different* perspectives on what time is, each arising from one of the three pillars of physics.

Firstly, according to special relativity, time is not absolute; it does not tick by independently of events taking place in three-dimensional space but must instead be combined with space into four-dimensional spacetime. This is not just a mathematical trick. It is forced upon us by the properties of the real world, tested again and

again in experiments and shown to be just the way the universe is. Einstein's theory of gravity (general relativity) then tells us that spacetime *is* the gravitational field itself—the stronger the field, the more curved spacetime is. So, the lesson from relativity is this: time is part of the physical fabric of the universe, a dimension that can be stretched and warped by gravity.

This is very different from the almost trivial role played by time in quantum mechanics, where it is nothing more than a parameter: a number that you plug into an equation. Knowing the state of a system at some time t_1 allows us to compute the state of the system at any other time, t_2, and so on. And it works in reverse too: knowing the state of a system at a later time t_2 allows us to compute it at an earlier time, t_1. The arrow of time in quantum mechanics is reversible.

In thermodynamics, time has yet another meaning. Here, it is neither a parameter nor a dimension, but an irreversible arrow pointing from past to future, in the direction of increasing entropy.

Many physicists believe that we will one day combine all these three different notions of time. For example, we have not yet heard the last word on quantum mechanics, since we still do not fully understand how the deterministic equations that describe the dynamics of a quantum state—and in which time can flow in either direction—tie in with the irreversible, one-way process of measurement. There are strong hints from the rapidly developing field of quantum information theory that the way a quantum system interacts and becomes entangled with its surrounding environment is similar to the way a hot object leaks heat to its colder surroundings. This would see a coming together of quantum mechanics and thermodynamics.

A neat experiment at the University of Queensland in Australia in 2018 showed just how puzzling this all is by demonstrating that at the quantum level, events occur with no definite causal order. Basically, in physics, causality means that if an event A takes place before an event B (in some frame of reference) then A may or may not have influenced or even caused B. But the later

event B could not have influenced or caused event A. At the quantum level, this sensible causality was shown to break down. This has led some physicists to argue that the arrow of time really does not exist at the quantum level but is only an emergent property when we zoom out to the macroscale.

However, it has been the quest to bring together the first two pillars of physics that has occupied the minds of so many physicists for a century. Entire careers have been dedicated to trying to understand how to combine quantum mechanics and general relativity into one all-encompassing theory of quantum gravity. This unification of the two most important ideas in twentieth-century physics is the subject of the next chapter.

UNIFICATION

Physicists' relentless drive to unify their theories—to bring together the laws of the universe and encapsulate them in a single neat mathematical equation—a 'theory of everything'—often appears to be no more than an obsession with simplicity and compactness, an effort to package up the complexity of all natural phenomena using the minimum number of underlying principles. In fact, it's subtler than that. Throughout the history of physics, the more we've discovered about the workings of the nature, the more connections we've found between seemingly unconnected forces and particles, and the fewer rules and principles we've needed to explain an ever-wider range of phenomena. Unification is not something we deliberately set out to achieve; it has emerged as a result of our deeper understanding of the

physical world. But this success undeniably comes with a certain aesthetic appeal that drives us to keep going along the same lines. And we have been astonishingly successful at it.

Mathematically, the quest to unify the laws of physics has often involved a search for abstract symmetries, patterns that hide deep truths about nature. We saw in chapter 2 just how central symmetry has proven to be in physics, and the way it leads to laws like the conservation of energy and momentum. But I'm afraid that to truly appreciate its importance and the role that different symmetries have played in theoretical physics over the past century is somewhat beyond the remit of this short book.

The hunt for a unified theory is sometimes described as an attempt to gather all the forces of nature into one framework, suggesting that there exists just one 'superforce' and that the different interactions we know of in nature—electromagnetism, gravitation and two short-range forces within the confines of atomic nuclei—are all different aspects of this single force. Physicists have so far had a good deal of

success with this broad project of unification. I have already described how Newton understood that what causes the apple to fall from the tree is the same universal force (gravity) that controls the motion of the heavenly bodies across the sky. This was not at all obvious at the time, even though it might seem so to us today. Before Newton, it was believed that objects fell to the ground because everything had a 'tendency' to move to its 'natural' place—towards the centre of the world—and that the motion of the Sun, Moon, planets, and stars was subject to very different principles. Newton's law of universal gravitation brings these phenomena together by stating that all masses are attracted towards each other, with a force proportional to the product of their mass and inversely proportional to the square of the distance between them. It doesn't matter whether it is an apple or the Moon; the same formula governs the way both are attracted by the Earth.

Another huge leap forward along the path to unification took place almost two centuries after Newton, when James Clerk Maxwell

showed that electricity and magnetism are in fact different facets of the same electromagnetic force. So, the electrostatic attraction between a scrap of paper and a balloon that has been rubbed on your clothing has its origin in the same electromagnetic force that attracts a paper clip towards a magnet. Almost all phenomena we see in nature are due ultimately to one or other of these two forces: gravity and electromagnetism. It was therefore natural to ask whether we can go further and bring them together in a combined theory.

We have already seen that, at a fundamental level, the gravitational field is nothing more than the shape of spacetime itself, a revelation that was also due to a unifying idea. By combining space with time, Einstein revealed a profound truth: that only in four-dimensional spacetime can all observers (however fast they are moving relative to each other) agree on the separation between two events. A decade later, his general theory of relativity gave the world a new and more accurate picture of how mass and energy cause this spacetime to curve. But that wasn't

enough for Einstein, who spent the most part of the next four decades of his life searching unsuccessfully for a unified theory that would combine his theory of gravity with Maxwell's electromagnetic theory.

We now know that there are, in addition to gravity and electromagnetism, two other forces—the strong and weak nuclear forces—that only act over very tiny distances, but which are just as important as far as the fundamental laws of nature are concerned. And it was the unification of the electromagnetic force with one of these nuclear forces that would be the next step forward in twentieth-century physics.

But this important advance in our understanding of the nature of the fundamental forces only came about with the evolution of quantum mechanics from a theory describing the microcosm in terms of particles and waves to one involving fields. I touched very briefly on the meaning of fields in chapter 3 in the context of gravity and electromagnetism. We are now ready to tackle the meaning of a quantum field.

QUANTUM FIELD THEORY

I may have given you the impression that once quantum mechanics was completed almost a hundred years ago, most physicists busied themselves with applying it to real problems in physics and chemistry, leaving just a few of the more philosophically minded to carry on arguing about what it all meant. To a large extent, this version of history is true. But it is also true that quantum mechanics continued to develop in sophistication throughout the first half of the last century. The basic mathematical formalism—the equations and the rules—was certainly in place by the late 1920s, but Paul Dirac soon managed to combine quantum theory with Einstein's special theory of relativity. He also brought together quantum mechanics and Maxwell's electromagnetic field theory to produce the very first quantum field theory. This developed into a powerful and very precise way of describing the electromagnetic interaction of matter with light at the quantum level.

Dirac's quantum field theory describes how electrons emit and absorb photons, and how two

electrons will repel each other, not by some invisible force that links them across space, but by the exchange of photons. By the 1930s, the distinction, at the quantum level, between the physics of particles and the physics of fields was swept away. So, in the same way that photons are the particle-like manifestations of the electromagnetic field—lumps of pure energy at the quantum scale—so too are the localised particles of matter, such as electrons and quarks, just manifestations of their more fundamental associated quantum fields. However, unlike photons and the electromagnetic field, this is not so obvious when it comes to the matter particles. The reason for this is that photons can bunch together in unlimited numbers, giving rise to what we perceive as an electromagnetic field on the macroscale, whereas matter particles like electrons and quarks are less sociable, thanks to one of the rules of quantum mechanics called the Pauli exclusion principle, which states that no two identical matter particles can occupy the same quantum state. This means we do not perceive their quantum fields so easily.

By the late 1940s, mathematical problems with the description of quantum fields were finally resolved, and the theory known as quantum electrodynamics (QED) was completed. To this day, it is regarded as the most accurate theory in all of science. It is also the physical theory that explains at a fundamental level almost everything in the world around us, since it underpins all of chemistry and the nature of matter—from the way the electronic circuitry and microchips in my laptop work to the neurons firing in my brain, commanding my fingers to move across the keyboard. This is because QED is at the heart of all interactions between atoms.

And yet, for all its power, QED still describes only *one* of the four forces of nature: electromagnetism.

During the late 1950s and '60s, physicists used beautiful but complicated mathematical reasoning to combine QED with a field theory of the weak nuclear force. They showed that the weak force was, at a fundamental level, also generated by exchanged particles equivalent to the role played by exchanged photons in describing the

electromagnetic force. Today, we have a unified theory describing a single 'electroweak' interaction that, through a process called symmetry breaking, splits into two distinct physical forces: electromagnetism (manifested by the exchange of photons) and the weak force, carried by the exchange of the W and Z bosons, which were subsequently discovered at CERN in 1983 and have since been extensively studied. The split between the two forces (the symmetry breaking) is due to another field, called the Higgs field, which gives the W and Z particles mass while leaving the photon massless. This unification means that, at a fundamental level, the four forces of nature are reduced to just three: the electroweak force, the strong nuclear force, and gravity (which in any case is not actually a force at all, according to general relativity). You may disagree with me as to whether this has helped simplify matters.

In parallel with this advance, another quantum field theory was developed to describe the strong nuclear force that holds the quarks together inside protons and neutrons. A subtlety

of the strong force is that the way it acts between quarks involves a property called 'colour charge', which deserves a brief mention. Just as particles that feel the electromagnetic force come in two types of electric charge, which we refer to simply as positive and negative,[1] the particles that feel the strong force (quarks) come in three types of 'charge', named colour charge to distinguish them from electric charge. Note that the word 'colour' here is not to be taken in any way literally. The reason three types of colour charge were needed, rather than just two (as with electric charge), was to explain why protons and neutrons must each contain three quarks; and the reason the analogy with colour was chosen was because of the connection with the way the three different colours of light (red, blue and green) combine to produce white light. Thus, the three quarks in a proton or a neutron each carry a different colour charge: red, blue or

---

1 And which could equally well have been called 'left' and 'right', 'black' and 'white', or 'yin' and 'yang', to indicate that they are opposite to each other.

green, which combine to produce a particle that has to be 'colourless'.

The rule was that quarks could not exist by themselves because they carried colour; they could only exist in nature by sticking together to make up colourless combinations.[2] For this reason, the field theory of the strong nuclear force that binds quarks together became known as quantum chromodynamics, or QCD. The exchange particles between quarks are the gluons, a rather more evocative and appropriate name, I think you'll agree, than that of those weak force–carrier particles, the W and Z bosons.

Let us then take stock. Of the four known forces of nature, three are described by quan-

---

2 The other type of particles made of quarks, called mesons, contain a quark and an antiquark, which must both have the same colour charge because antiparticles always carry the opposite properties. So, you could have a meson made up of a red quark (of some flavour, such as up, down, or strange) together with an anti-red quark of some other flavour. The flavours of the quark and antiquark define the type of meson, while their colour and anti-colour cancel out to ensure a colourless particle. Complicated? You betcha!

tum field theories. The electromagnetic and the weak nuclear force are linked together by the electroweak theory, while the strong force is described by quantum chromodynamics. A yet-to-be fully developed theory that connects these three forces together is known as a grand unified theory (or GUT). But until we find one, we must make do with a loose alliance of the electroweak theory and QCD, known as the Standard Model of particle physics.

Even its most ardent defender will admit that the Standard Model is probably not the last word on the matter. It has survived this long in part because we have nothing better to replace it with and in part because its predictions have so far been validated by experiments, such as the discovery of the Higgs boson in 2012 (more of which later). And yet despite this being the best description we have of three of the four forces of nature, physicists would like nothing better than make some new discovery that conflicts with the Standard Model, in the hope of discovering a deeper and more accurate description of reality. But as long as

the predictions of the Standard Model continue to be confirmed by experiments, it lives to fight another day.

Of course, this whole discussion of quantum field theories omits a very important ingredient: gravity.

THE QUEST FOR QUANTUM GRAVITY

We have discovered that the description of our everyday world at the length, time, and energy scales appropriate for Newtonian physics is only an approximation, and that beneath it are more-fundamental physical theories that come into their own at the extreme scales. At one end, we have quantum field theory, which has led us to the Standard Model of particle physics and which accounts for three of the four known forces in the universe. At the other extreme, we have the general theory of relativity, which gives us the Standard Model of cosmology that encompasses the other force, gravity. This standard model of the very large is called a variety of different names, such as the concordance model,

or the Lambda–cold dark matter model, or the Big Bang cosmology model. I will discuss it more fully in the next chapter.

Therefore, a question that physicists are often asked is *why* we feel it is so important, indeed whether it should even be possible, to keep going with our obsession with unification, to try to combine these two models describing entirely different scales: the quantum realm and the cosmic realm. Surely each works well in its own domain, and that should be enough for us. But again, I must stress that the purpose of physics is not simply to account for what we observe or to find some useful application based on it; physics is about gaining the deepest and most complete understanding of reality.

So, this is where we are at the moment: stuck with two successful frameworks—quantum field theory and general relativity—which just don't seem to want to fit together. Indeed, they appear to have very little in common: their mathematical structures are incompatible. And yet this cannot be the whole story. We know that spacetime reacts to the matter

within it. We also know that matter at the subatomic scale behaves according to the rules of quantum mechanics, which must surely in turn affect the behaviour of spacetime. If an unobserved electron is in a quantum superposition of being in two or more states at once, as we know electrons can be—for example, if their quantum state is spread out over some volume of space or in a superposition of different energies at once—then surely the spacetime around this electron must reflect this fuzziness too. The problem is that general relativity just *isn't* 'quantum-y', and it is far from straightforward how we can make it so. One of the problems with this is that subatomic particles have such tiny masses that their effect on spacetime is nigh impossible to measure.

Still, the issue remains: How do we quantise the gravitational field? What do we need to do to bring quantum field theory and general relativity together? And if they are truly as incompatible as they seem to be, then which one of these two incredibly successful theories needs to 'give way' to get us to quantum gravity?

STRING THEORY

In the mid-1980s, a candidate theory of quantum gravity was developed. It was based on a mathematical idea called supersymmetry, which I mentioned briefly in chapter 2. This candidate theory became known as superstring theory, and it captured the imagination of many mathematical physicists of my generation. Supersymmetry suggests a relationship between the two general types of elementary particles in the Standard Model: the matter particles, or fermions (quarks and electrons and their cousins), and the force carrier particles, or bosons (the photon, gluons, and W and Z bosons).

String theory had originally been proposed in the late 1960s as a theory of the strong nuclear force, but when quantum chromodynamics was developed in the '70s and found to be so successful, string theory fell out of favour and was seen as no longer necessary. But it was soon realised that by incorporating into string theory the idea of supersymmetry, it could be reborn as a candidate for a much grander undertaking

than a theory of the strong force: a theory of everything.

The basic premise of supersymmetric string (or superstring) theory is that one way to unify all the forces is to add more dimensions to space beyond the three we are aware of. This idea goes back to work by the Polish theoretical physicist Theodor Kaluza, who noticed, just after the end of the First World War, that if he solved Einstein's field equations of general relativity in five-dimensional spacetime instead of four, then electromagnetism emerged out of the mathematics as vibrations in this fifth, unseen, dimension. Kaluza showed his work to Einstein, who initially liked it. It seemed to do for electromagnetism what Einstein had achieved for gravitation: changing its fundamental description from a physical force to pure geometry.

Yet, despite this elegant way of unifying light (electromagnetism) and gravity (general relativity), most physicists—including Einstein himself—soon became sceptical of Kaluza's work, as there was no experimental evidence

to suggest that this extra dimension of space existed.

A few years after Kaluza's original idea, the Swedish physicist Oskar Klein suggested that the reason the fifth dimension is hidden is because it is curled up on itself, and therefore too tiny to be detected. There is a standard analogy that helps to explain what this means. From a distance, a hose looks like a one-dimensional line, but zoom in and you see that it is in fact a two-dimensional surface wrapped around into a cylinder. The second spatial dimension (the circular direction around the hose) was too small to be seen from a distance. Klein suggested the same thing applied to Kaluza's fifth spatial dimension, which was curled up into a circle a billionth of a trillionth of the size of an atom. Despite Kaluza-Klein theory not leading to a unification of gravity and electromagnetism, it did help researchers understand the relevance of the higher dimensions in superstring theory. However, now, instead of just one hidden spatial dimension, there needed to be six, all rolled up into an impossible-to-visualise six-dimensional

ball. Superstring theory thus states that there are ten dimensions: four of spacetime that we experience plus the six hidden dimensions.

To this day, many researchers looking to unify the forces of nature still work on string theory. They argue that we have come so far, using successful ideas like quantum field theory and supersymmetry to understand three of the four forces; therefore surely gravity can also be tamed. They may well be right.

String theory begins with the quantum mechanical properties of matter within spacetime. Its central idea is that all elementary point-like particles are in fact tiny strings, vibrating in the hidden dimensions. These strings would be far smaller than the scales currently probed by particle physics and so we can only experience them as point particles. The problem that emerged by the 1990s was that it appeared there were five *different* versions of string theory, and no one knew which one was the correct one. So a new, even grander framework was proposed which unified all five versions under one umbrella. This all-encompassing framework is now called

M-theory, which is a supersymmetric theory with eleven, rather than ten, dimensions. Yet again, it seemed, another hidden dimension was needed to help with the grand unification programme.

So, is that it? Is M-theory our ultimate 'theory of everything'? Sadly, we cannot yet say. While the mathematics is elegant and powerful, we still don't know if string theory or M-theory are the right descriptions of reality. In the next chapter, I will discuss some of the outstanding issues and controversies surrounding this subject. In any case, M-theory has a worthy adversary in the race for unification. This rival theory is just as speculative, but a number of theoretical physicists see it as a purer and more sensible way of tackling unification. It is called loop quantum gravity, and it came to the fore in the last decade of the twentieth century.

LOOP QUANTUM GRAVITY

Loop quantum gravity does not start from quantum field theory, but from the other direction—from general relativity. It assumes that spacetime

itself, rather than the matter it contains, is the more fundamental concept. Aesthetically, it would seem sensible to try to quantise the gravitational field—which, according to general relativity, *is* spacetime itself. Thus, if we shrink down to small enough length scales, we should see space become grainy and discrete. In the same way that Max Planck proposed, in 1900, that heat radiation ultimately comes in quantum lumps, quantizing space suggests there should be a smallest length that cannot be further divided. However, the quanta of gravitational energy are the quanta of space itself, which means that they don't exist as lumps *within* space . . . they *are* lumps *of* space.

It is thought that the tiniest unit of space—a quantum of volume—is one Planck length, or 10^{-35} m, across. I have always enjoyed trying to find ways of describing how tiny this volume is. For example, an atomic nucleus contains as many Planck volumes inside it as there are cubic metres in the Milky Way galaxy.

This discretisation of space seems inevitable if we want to quantise the gravitational field. And it

therefore follows that time must also be 'lumpy'. So the smooth space and time that we experience is nothing more than a large-scale approximation of the lumpy quanta of gravity, smeared out because the individual pixels of spacetime are too small for us to perceive.

Loop quantum gravity contrasts dramatically with string theory, which predicts that, just as the three forces covered by the Standard Model (electromagnetism and the strong and weak nuclear forces) are in fact quantum fields manifested as force-carrying particles, so too is the gravitational field mediated by a quantum particle of gravity: the graviton, a massless state of a string. In string theory, this quantum of the gravitational field exists *within* spacetime, whereas in loop quantum gravity, it is spacetime itself that is quantised.

Loop quantum gravity refers to the closed paths that take you from a quantum of space, via its links to adjacent quanta, around in a loop and back to the starting point. The nature of these loops determines the curvature of spacetime. They are not physical entities like strings.

The roads to Unification
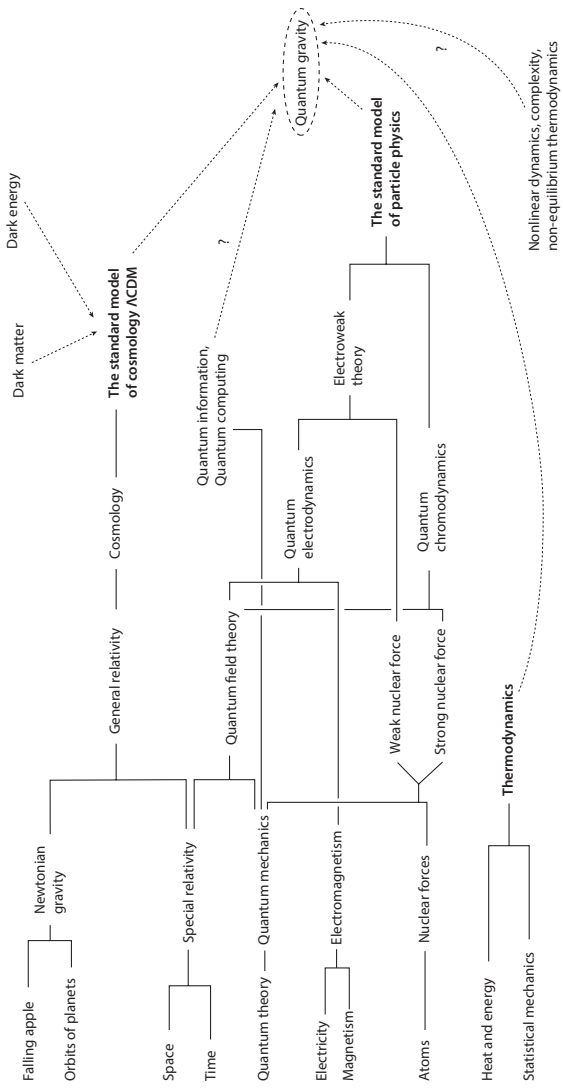
All that is real is the relationship between the loops.

In a sense, loop quantum gravity is modest in its scope. But when you consider it more carefully you begin to realise that, if it is indeed the correct description of reality, then it is not so much that events take place within space and for a duration of time, but rather that the universe, and everything in it—all matter and energy—is nothing more than quantum fields coexisting and superimposed onto each other. And these fields do not require space and time to exist in, since spacetime is itself *one* of these quantum fields.

In summary, we cannot yet claim to have a genuine theory of everything, nor do we understand yet how to bring quantum mechanics and general relativity together. Rather, we have

FIGURE 4. (OPPOSITE) Unification—A (simplified) chart showing how concepts in physics (theories, phenomena, forces) have come together over the years. Note that while the chronology is correct (running from left to right), you should not read too much into it. For example, special relativity appears directly below Newtonian gravity even though it came centuries later.

candidate theories that show some promise, but which still leave many questions unanswered. Brilliant physicists have built their careers on one or the other such theory, but just as with different interpretations of quantum mechanics, there is a lot of sociology of science involved, and views on which theory shows the most promise really do depend on who you speak to. Broadly then, in the red corner we have string theory, which is our current best stab at unifying the four forces of nature, but which despite three and a half decades of research is still speculative. Some physicists claim that despite all the progress it has made, it is now reaching something of a crisis because it hasn't delivered on its early promise. Indeed, it can be argued that it is not yet even a proper scientific theory, since it has not made any testable predictions. Then, in the blue corner, we have loop quantum gravity, which seems to be the most sensible way of quantising spacetime, but which does not tell us how to then combine gravity with the other three forces. Whether one or the other of these two approaches is correct, or whether we need

to somehow merge the two, or even look for an entirely new theory, we still do not know.

This leads us nicely on to where the current outstanding problems and controversies are in fundamental physics, and what advances are likely in the coming decades.

THE FUTURE
OF PHYSICS

The remarkable success of twentieth-century physics might suggest that all we have left to do is iron out a few creases, refine our experimental measurements, and put the finishing touches to our mathematical theories—that most of what there is to know is already known and we are just dotting a few 'i's and crossing the final 't's. You may have got the impression that there is no need for another Newton or Einstein (or indeed a Maxwell, Rutherford, Bohr, Dirac, Feynman, Witten, or Hawking) to come along and bring about a new revolution in physics, because we are already within touching distance of a theory of everything that explains the workings of the entire universe.

Unfortunately, or fortunately if you're a research physicist just starting out in your career

and looking for big problems to tackle, the truth is far from this. In fact, I would say we are further away from the end of physics today than we thought we were twenty or thirty years ago. We speak of the Standard Model as describing all the elementary building blocks of matter and energy, but we are now pretty sure that everything we have found only makes up 5 percent of the universe. The other 95 percent, known as dark matter and dark energy, is still to some extent mysterious. We are confident it's out there, but we don't know what it is made of or how it fits into our current theories. In this chapter, I will explore this mystery along with other outstanding challenges in fundamental physics.

DARK MATTER

The rotational speed of galaxies, the motion of entire galaxies within galaxy clusters, as well as the large-scale structure of the entire universe, all point towards a significant component of the universe consisting of a near-invisible matter component. We call it 'dark,' not because it is

hidden behind other, visible matter, or even because it is actually dark, but because, as far as we can tell, it doesn't feel the electromagnetic force and so does not give off light or interact with normal matter, other than gravitationally,[1] and so a better name for it would have been in-visible matter. Think for a moment about why, if you slam your hand down on a solid table, it doesn't pass straight through. You might regard this as trivial: surely it is because both your hand and the table are made of solid stuff. But don't forget that down at the level of atoms, matter is mostly empty space—diffuse clouds of electrons surrounding a tiny nucleus—and so there should be plenty of room for the atoms that make up

---

1 Of course, there are matter particles, such as neutrinos, that also do not feel the electromagnetic force. However, they interact instead with other matter via the weak nuclear force, and so are not what is referred to as dark matter. Even dark matter itself may yet be found to interact via one or more of the other three forces, but it would have to be very weakly indeed (or we would have measured it by now). Physicists haven't entirely given up hope of such a small non-gravitational interaction, as that would increase the chances of dark matter particles being detected or created in an accelerator.

your hand to easily pass through the atoms of the table without any physical matter coming into contact. The reason they don't is because of the electromagnetic force between the electrons in the atoms of your hand and the electrons in the atoms of the table, repelling each other and providing the resistance we experience as solidity. However, if your hand were made of dark matter, then it would pass straight through as though the table weren't there—the gravitational force between them being too weak to have much effect.

It has long been known that galaxies have much more mass than can be accounted for if one measures all the normal matter they contain in the form of stars, planets, and interstellar dust and gas. At one point it was thought that dark matter might be made up of long-dead stars and black holes—objects made of normal matter, but which do not emit light. However, overwhelming evidence now suggests that this invisible stuff must be made up of a new form of matter, most likely a new type of particle yet to be discovered.

Originally, dark matter was proposed to explain the large-scale dynamics of entire clusters

of galaxies. Further evidence then came from the way stars moved within spiral galaxies, circulating like undissolved coffee granules on the surface of a mug of instant coffee after it has been stirred. Most of the stars—and hence, you would think, most of the mass—in a galaxy are concentrated at its core, which would require those in the outer rim to be moving around the centre more slowly. The observed higher-than-expected orbital speeds of these outer stars suggest that there must be some additional invisible stuff present, extending out beyond the visible matter we can see and providing the extra gravitational glue to stop the outer stars from flying off.

Dark matter can also be seen from the way it curves space around it. This phenomenon manifests itself in the way light bends while on its path from very distant objects to our telescopes. The amount of bending can only be explained by the extra gravitational curvature of space provided by the dark matter of galaxies that the light passes on its way to us.

So, what do we know about dark matter other than that it provides this necessary extra gravi-

tational attraction? Might this not be accounted for by something less exotic than a new form of matter? Indeed, some astrophysicists suggest that there may be no need for dark matter at all—if we are allowed to modify the properties of the gravitational force at large distances. An idea known as MOND (modified Newtonian dynamics) is one such suggestion and, on the face of it, this idea can sound quite appealing. However, while MOND, or other related hypotheses that modify general relativity, can explain away some of the observed effects, there is plenty that they do *not* explain. None of these models have been able to match the observational data for galaxy clusters, particularly colliding galaxy clusters (such as the famous Bullet Cluster), the detailed structure of the cosmic microwave background radiation, globular star clusters, or, more recently, tiny dwarf galaxies.

The existence of dark matter also seems necessary to explain the structure of the early universe. In contrast with normal matter, which through its interaction with the electromagnetic field kept its energy high, dark matter cooled

down more quickly as the universe expanded and therefore started to clump together gravitationally earlier. One of the most important results in astrophysics in recent years has been the confirmation from sophisticated computer simulations of galaxy formation that we can only explain the real universe if it does indeed contain large quantities of dark matter. Without it, we would not get the rich cosmic structures we see today. Put more bluntly, without dark matter, most galaxies, and hence stars and planets, could never have formed in the first place. This remarkable conclusion is supported beautifully by data showing subtle fluctuations in the temperature of deep space, the imprint of the very young universe on the cosmic microwave background radiation. It was recognised back in the late 1970s that these fluctuations in the cosmic microwave background, while helpful in providing the seeding for the present-day distribution of matter in the universe, were too tiny to explain how galaxies could form. Dark matter helped provide the extra clumping that was needed. It was one of the great scientific triumphs of the

end of the twentieth century when the COBE satellite[2] measured these fluctuations to be just what had been predicted. Since then, further space missions have mapped these wrinkles in the cosmic microwave background with ever-increasing resolution: NASA's WMAP mission in the first decade of this century, then the European Space Agency's Planck satellite, which launched in 2009.

While we are left in little doubt that dark matter is real, we are still in the dark as to what it is made of. It is a growing source of frustration in astrophysics that, in parallel with the accumulation of evidence in support of dark matter, we have failed to find out what it actually is. The consensus now is that it consists of a new type of heavy particle (heavy by the standards of elementary particles, that is), and most of the experimental effort thus far has been

2 The Cosmic Background Explorer, also referred to as Explorer 66, was a satellite dedicated to cosmology, which operated from 1989 to 1993. Its goals were to investigate the cosmic microwave background radiation of the universe.

focused on building sophisticated underground detectors that can capture extremely rare events when such a dark matter particle streaming in from space collides head-on with an atom in the detector. To date, no signal from these increasingly sophisticated and sensitive experiments has been picked up.

And yet, physicists looking for dark matter remain optimistic. Most likely, they say, it will be in the form of slow-moving particles, making up what is known as 'cold dark matter'. And there is no shortage of suggestions for what these particles might be, with such wonderful-sounding names as axions, sterile neutrinos, WIMPs,[3] and GIMPs[4]. Many feel confident that experimental evidence will emerge soon. But then, we've been saying that for some time now.

I should at this point say just a little about neutrinos, which for a while were the leading candidates for dark matter. These are elusive yet abundant particles that we know exist, which have a

---

3 Weakly interacting massive particles

4 Gravitationally interacting massive particles

tiny mass and are almost invisible. You would need a light-year's thickness of lead shielding to have even a fifty-fifty chance of blocking them. You could say that they are, to all intents and purposes, 'dark matter'. However, they cannot be *the* dark matter that we are searching for because, being so light, they travel at near light speed—too fast to remain bound within galaxies and thus to explain galaxies' anomalous properties. We refer to neutrinos as hot dark matter, because they move so fast.

And as if the unresolved problem of dark matter weren't big enough for physicists, we now know of another mysterious substance filling the universe, which plays a vital role in shaping it.

DARK ENERGY

In 1998, astronomers studying the faint light of supernovae in distant galaxies used it to calculate the speeds at which those galaxies were receding from us due to the expansion of the universe. They found that they were moving away more slowly than their distance from us suggested

they should be. Since the light now reaching us from these galaxies left them when the universe was very young, their slower-than-expected recession speeds meant that the universe must have been expanding more slowly in the past. So, rather than the cumulative gravitational attraction of all the matter in the universe—both normal and dark matter—slowing down the expansion of the universe, something else was at work, making it expand more quickly now than it did in the past.

This mysterious repulsive substance acting against gravity and stretching space ever more quickly became known as dark energy. According to our present understanding, dark energy may ultimately result in what is called the 'heat death' of the universe many billions of years from now as space continues to expand ever more rapidly and to cool as it settles towards a state of thermodynamic equilibrium. But until we truly understand the nature of dark energy, and indeed the properties of the very early universe (see the next section), we should not be too quick to conjecture about its final fate. It's

a long way off, and anything could happen between now and then!

Until a few years ago, I would have said we know even less about dark energy than we do about dark matter, but that is now changing. There is a quantity in Einstein's equations of general relativity, known as the cosmological constant (and denoted by the Greek letter Λ, or lambda), that fits the bill. What we call dark energy is most likely the energy of empty space itself—what is referred to as the quantum vacuum. We have seen how everything ultimately comes down to quantum fields in the end. All the different particles that make up matter and energy, whether quarks, electrons, photons, or Higgs bosons, can be regarded merely as localised *excitations* of these quantum fields—like waves on the surface of an ocean. However, if you were to remove all the particles from a volume of space, this does not get rid of the field. Instead, we say it is left in its ground, or vacuum, state, but there will still be virtual particles popping in and out of existence within this vacuum all the time, borrowing the energy from their surroundings in

order to exist, but paying it back just as quickly when they disappear again. So to say that the quantum vacuum of empty space has zero energy would be the same as claiming a calm ocean has no depth. The equivalent of water beneath the ocean surface is this dark energy—it is the cosmological constant.

However, having a mathematical symbol for dark energy does *not* mean we entirely understand its nature yet. Astronomical measurements suggest that the cosmological constant has a certain numerical value, but, like the mass of the Higgs boson in the Standard Model, we do not know why it has this value. This long-standing problem in physics is known as fine-tuning and is very unsatisfying. In fact things are even worse than this. The discrepancy between this calculated vacuum energy from quantum field theory and the observed vacuum energy from cosmological measurements is so huge that it's one of the most embarrassing and unresolved problems in physics. You see, the calculated value is a ridiculous 120 orders of magnitude bigger than the observed value.

Our 'best guess' cosmological model—the equivalent of the Standard Model in particle physics—which contains under its umbrella what we currently know about dark matter and dark energy, is called the ΛCDM model (or Lambda–cold dark matter model). And similar to the way deeper quantum field theories underpin the loose alliance of the Standard Model of particle physics, so too does general relativity underpin the ΛCDM cosmological model.

There is one more important ingredient of the ΛCDM model, which most, but by no means all, cosmologists claim is needed to explain the properties of the universe we see. It is called cosmic inflation, and it provides a possible answer to that perennial question: How did the universe and all the matter and energy it contains come into being in the first place?

INFLATION AND THE MULTIVERSE

As we touched on at the very beginning of this book, since the dawn of human history we've created many myths about the origins of the

universe. Today, physics has given us a demystified explanation of how the universe began, with overwhelming observational evidence to back it up. But did the Big Bang itself have a cause? Was there something that triggered the birth of our universe in the first place?

The simplest answer is that there was no 'before' the Big Bang, for it marked the birth of both space and time. An idea put forward by Stephen Hawking and James Hartle, called the 'no boundary' proposal, states that, as we wind back the clock closer and closer to the Big Bang, time begins to lose its meaning and becomes more like a dimension of space. We therefore end up with smooth four-dimensional space at a point of the universe's origin. So it is meaningless to ask what happened before the Big Bang, in the same way that it is meaningless to ask what point on the surface of the Earth lies south of the South Pole.

But the Big Bang model is not enough on its own to explain the universe we see today. In particular, two problems puzzled cosmologists half a century ago. The first is called the flatness problem. This is yet another fine-tuning issue and re-

lates to the density of matter and energy in the universe, which appears to have a just the right value to make space almost perfectly flat.[5] The second problem is called the horizon problem. The furthest we can see out into space is probably only a tiny fraction of the entire universe, and there exists a horizon beyond which we can never see. This horizon marks the edge of what is known as the visible universe. It exists because the universe hasn't been around forever, and light takes a certain time to reach us. An added complication is that the universe is expanding, and at some distance space is stretching faster than light can travel through it (like trying to walk up a rapidly descending escalator).

Consider a galaxy near the edge of the visible universe in one direction and another galaxy near the edge in the opposite direction. Due to the expansion of the universe, any intelligent

---

5 It is hard to visualise what we mean by 'flat' 3-D space. The easiest way is to imagine restricting our space to just two dimensions. Now, it is clear that a page in a book is flat, whereas the surface of a ball is not.

beings living in one of these distant galaxies would be completely unaware of the existence of the other galaxy, since light from it won't have reached them yet, nor will it ever. In fact, the regions of space containing the two galaxies could never have been in contact and cannot ever have shared information. Why is this a problem? Because in every direction we look and as far out as we can see, the universe looks the same. Both of those distant galaxies look very much the same (to us in between them) in terms of their physical properties, composition, and the structure of matter within them. How can this be if they were never in contact in the past?

To solve these two issues—the flatness problem and the horizon problem—a concept known as cosmic inflation was proposed forty years ago. It went like this: when the universe was just a fraction of a second old, it underwent a short period of exponential expansion due to yet another quantum field, called the inflaton field, during which it grew stupendously rapidly to a trillion trillion trillion trillion times the size it was before. This solves the problem of the finely tuned

density giving rise to the flat spacetime we see today, because any tiny amount of curvature got stretched out by inflation.

The way inflation solves the horizon problem is more interesting. The usual explanation is that distant parts of the universe that do not appear to ever have had the chance to be in contact, and thus to synchronise their physical properties, were in fact in contact at the beginning, but that inflation caused space to expand so rapidly they appear now to be too far apart to have ever been causally connected.

I said this is the 'usual explanation', but if you think about it, there are two things not quite right with referring to inflation as being 'rapid' expansion. First, for distant parts of the universe to be able to communicate with each other when they were close together surely required them to stay closer together for longer, not zoom apart too quickly. Secondly, when we refer in mathematics to something being exponential, we mean that it varies slowly to begin with and then speeds up (the slope becoming steeper). This is a better way of thinking about

the inflationary early universe. It started expanding slowly, then sped up. Then, at some point, this exponential expansion changed to what is called a 'power law' expansion, where instead of the expansion speeding up, it started to slow down again—until, that is, dark energy kicked in halfway through the universe's life and started to speed the expansion up again.

Of course, this tells you nothing about why the idea is so attractive, or why or how it works. So, let us spend a little time unpacking its meaning.

To appreciate how inflation works, you must first understand the difference between positive and negative pressure. Imagine you are holding an inflated balloon. The air inside it exerts a pressure on its inner surface, pushing outwards. If you were to now squeeze the balloon between your hands you would be expending energy to compress its air to a smaller volume, increasing its density, and this energy gets stored in the balloon's air molecules. Now, consider the reverse process: relax your hands so that the balloon expands back to its original size and the air inside becomes less dense again. The energy stored

in its molecules must now also drop back to what it was originally.[6] So, allowing the volume inside the balloon to expand means its energy decreases. This is the situation with 'normal', positive pressure: it loses energy as it expands.

But what if the balloon were filled with an unusual substance that did the opposite? What if, when it expanded in volume, its density didn't drop, but remained constant, and the energy per unit volume stayed the same, too, so its total energy *increased*? This is what we mean by something with 'negative pressure'—instead of the energy of the balloon's contents increasing when compressed, it increases when it *expands*. The closest example to this in our everyday world is a rubber band, since stretching it puts more energy into it.

And this is exactly what we have with the inflaton field filling space: it's like a rubber band, and has the property that every time the

---

6 Of course, this energy does not return to the muscles in your arms, but is lost instead as waste heat to the balloon's surroundings.

volume of space doubles, its total energy also doubles in order to maintain a constant field density. So, the inflaton field gives the universe energy, just as you would give a rubber band energy by stretching it.

You should be asking two questions here. Firstly, why does the inflaton field cause the expansion of space? After all, a rubber band does not spontaneously expand of its own accord. And secondly, given that the inflaton field generates energy, where does that energy originate from? Both questions have crafty but logical answers that can be found, you won't be surprised to hear, in the equations of general relativity.

Einstein's field equations tell us that gravity can be caused by pressure as well as by mass and energy. So, while something with positive pressure, like the molecules of air in a balloon, gives rise to normal attractive gravity, a substance with negative pressure will cause the opposite: antigravity—pushing everything apart rather than pulling it together. The inflaton field has the property that the repulsive effect of its negative pressure (or antigravity) is bigger than the nor-

mal attractive gravity caused by its energy, and so it causes space to expand at an accelerating pace.

As for where the energy of the inflaton field came from in the first place, the answer is that it is borrowed from its own gravitational field. Think of a ball on top of a hill: it has a store of positive potential energy that it can convert into kinetic energy if it rolls down. But a ball at the bottom of the hill has no potential energy, while a ball down in a hole has negative potential energy (since it needs energy put into to lift it back up to ground level). It is as though our universe began with no space and no energy, but a quantum fluctuation caused it to start rolling down a gravitational energy slope. As it rolled down it gained positive energy, paid for as it descended deeper into the gravitational valley by its increasing negative gravitational potential energy (see figure 5). Cosmologist refer to this as the ultimate free lunch—something from nothing. It is a very neat answer to the question: where did all the matter and energy in the universe come from in the first place?

Another way of understanding why gravitational energy is negative is to consider the following example: Start with two masses infinitely far apart, with zero gravitation energy between them. As they drift together they will gradually gain in gravitational attraction, but this gravitational energy is negative in the sense that you would need to put positive energy in to pull them apart again and get them back to the zero energy they started with.

When inflation ended, the energy of the inflaton field decayed into normal energy, which condensed out into all the matter we have today. The stuff of the universe was created from energy borrowed from its own gravitational field—the ultimate in creative accounting.

But just because inflation theory solves these problems in cosmology does not mean that it is correct. While most cosmologists subscribe to this theory, there are others who disagree, and there are indeed some subtle issues that have not yet been resolved. One critic is Stephen Hawking's long-term collaborator Roger Penrose. Instead of inflation, Penrose has proposed his
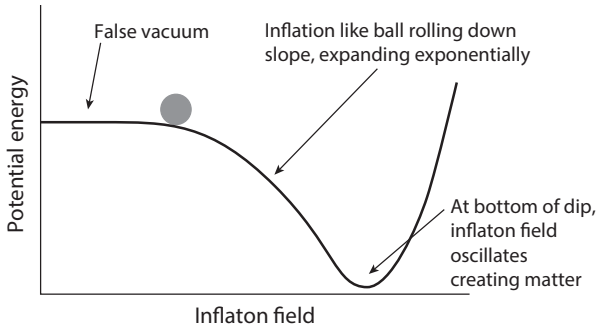
FIGURE 5. Inflation—The universe gained its positive energy (from which it created all matter) by 'rolling down' a gravitational potential energy slope, expanding as it did so.

own model, called conformal cyclic cosmology, in which the universe goes through an infinite series of epochs, each of which starts with a phase resembling a Big Bang. At the end of each cycle, all that remains, after even the black holes have evaporated, is thermal radiation. This, Penrose conjectures, is similar to the smooth high-energy radiation that would fill the universe just after the Big Bang, and with a clever connection between the low entropy of the early universe and the high entropy at the end (nothing can escape the second law of thermodynamics), he

can attach the end of one aeon to the beginning of another and see everything start again with a new Big Bang. Suffice it to say that this proposal is even more controversial than inflation theory.

And since we are deeply in the realm of speculation, why stop now? A fashionable idea in cosmology at the moment is known as eternal inflation. In this scenario, our universe is just a small bubble within an infinite, higher-dimensional space known as the multiverse, which has been undergoing inflation forever. In this scenario, the Big Bang that created our universe was but a quantum fluctuation 13.82 billion years ago, which created a bubble in this eternally inflating space. The space within this bubble—our universe—stopped inflating and slowed down to expand at a more sedate rate, while the multiverse outside continued its runaway inflation. And so, rather than a very brief period of inflation taking place *after* the Big Bang, now we have things the other way around with *our* Big Bang marking the *end* of inflation in our part of the multiverse.

What's more, eternal inflation predicts there will be other bubble universes within the multiverse, possibly an infinite number of them, all forever separated from each other and all being driven apart rapidly by the ever-expanding inflaton field.

This idea has an added benefit that many cosmologists find appealing. I have mentioned before that physicists do not like fine-tuning—that is, for there to be no underlying reason for why certain physical quantities have the values they do. This comes to a head when we consider that our most fundamental constants have just the right values for a universe such as ours to exist. If gravity were ever so slightly weaker, galaxies and stars might never have formed, and if the charge on the electron were very slightly stronger, atoms would collapse, and complex matter could not exist. So, the eternal inflation multiverse theory answers the question: Why is our universe so finely tuned as to be suitable for stars and planets, and life, to exist? The answer is that all possible bubble universes can exist, all obeying the same laws of physics, but each with its own set of fundamental physical constants. We

just happen to be in one that is just right for life to emerge and contemplate how lucky it is.

I should also just add here, to avoid any confusion: these bubble universes are *not* the same thing as the parallel realities of the multiverse (or many-worlds) interpretation of quantum mechanics, which are due to the different possible outcomes of measurement of the quantum world. The bubble universes in the eternal inflation theory are not parallel, overlapping realities, but completely independent of each other.

And before I move on, I want to add one more important point. We might wonder whether our universe is infinite in extent (even though we cannot see beyond our visible horizon), and it could well be. So how can infinite space fit inside a finite bubble floating in the multiverse alongside other bubble universes? The answer is rather strange: to us, on the inside, the universe could be infinite in extent, but finite in time. However, this is because we have a warped view of space and time from within our bubble. Viewed from 'the outside', our universe would appear finite in size but existing in endless time (see figure 6). It's
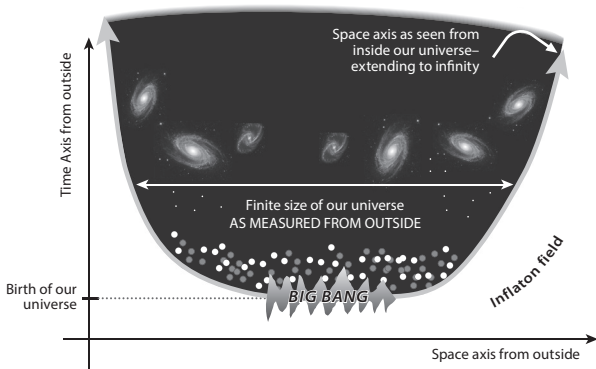
FIGURE 6. How can an infinite space fit into a finite volume?—Two views of our universe: from 'outside' it always has a finite volume, but for us, within our spacetime, the space axis is curved around so that it is pointing along the time axis to infinity. For us it looks as though all times coincide to give an infinite spatial extent.

a neat (but really conceptually tough—sorry!) way of understanding how infinite space can fit inside a finite volume.

INFORMATION

One topic I have not said much about that brings together all three pillars of fundamental physics—quantum mechanics, general relativity,

and thermodynamics—involves the role that information plays in physics. It is now understood that information is more than just an abstract notion and can in fact be quantified precisely. A long-standing puzzle, first highlighted by Stephen Hawking, was what happens to information, say this book you are reading, if you were to throw it into a black hole. The book will of course be lost forever, but what about the information it contained? By that, I mean the physical information that is encoded in the words of the book and that would be required to reconstruct them. You see, quantum mechanics tells us that information cannot be destroyed and must always be conserved.[7] Hawking described how black holes slowly evaporate, losing their energy in what is known as Hawking radiation, and quantum mechanics tells us that, in principle,

---

7 This comes about because, according to quantum mechanics, time is reversible. Therefore, just as a quantum state now uniquely determines a future state, so should a future quantum state uniquely determine a past one. But this wouldn't be possible if the information contained in this state is destroyed.

this radiation carries within it all the information that has been swallowed by the black hole, including the information needed to reconstruct this book. Do we know this for sure? Again, only a final theory of quantum gravity is going to lay this issue to rest.

The study of the mathematics of black holes has also led to the discovery that the maximum amount of information that can be stored in a volume of space is proportional, not to the volume of space, as might be expected, but to the surface area surrounding that volume. This idea became known as the holographic principle and is proving to be a powerful tool in theoretical physics. At its heart, it comes about because of a profound connection between information and energy. By storing more and more information in a volume of space, you increase its energy. And since energy is equivalent to mass, this means strengthening its gravitational field, to the point when the volume of space will collapse into a black hole. The holographic principle states that all the information will now be encoded on the black hole's event horizon. It is thought that this

idea even applies to the information needed to describe the entire universe. The role of information is likely to become increasingly important in connecting up the three pillars of physics.

ER = EPR

In 2013, two leading physicists, Juan Maldacena and Leonard Susskind, proposed an idea that might yet provide a new route toward unifying gravity and quantum mechanics. While it is far too early to judge if they might be right, it's also too fascinating for me not to mention it in passing. Known simply as ER = EPR, it suggests that there may be a deep and profound link between quantum entanglement (two particles connect across space) and wormholes in spacetime. But note that 'ER = EPR' is not an algebraic equation, despite the 'equals' sign (otherwise you might wish to cancel the E and the R from both sides, leaving just P = 1, which is meaningless). Instead, it refers to the initials of the authors (Einstein, Podolsky, and Rosen) of two classic papers published just a few weeks apart in 1935.

These two papers had hitherto been thought to be completely unrelated. 'ER' refers to Einstein and Nathan Rosen, who proposed that two black holes might be linked by a tunnel outside our dimensions, an idea that emerges from the mathematics of general relativity. 'EPR' refers to the second paper the two published with Boris Podolsky, in which they outlined their misgivings about the idea of entanglement in quantum mechanics—what Einstein referred to as 'spooky connections'. The novel suggestion of Maldacena and Susskind is that both these profound ideas, wormholes and entanglement, might in fact be one and same phenomenon. Time will tell if they are on the right track.

A CRISIS IN PHYSICS?

Will we then ever reach a complete understanding of reality, or will we be forever peeling back layers of the onion to reveal deeper truths underneath? This has certainly been true so far. First, we discovered that everything is made of atoms, then that these atoms are themselves made of

smaller parts—electrons orbiting around a dense nucleus. Later, we peered into the nucleus itself to discover that it is made up of smaller building blocks: protons and neutrons, which in turn are composed of even tinier quarks, which are themselves manifestations of energy fields—or, possibly, even far smaller vibrating strings in higher dimensions. Will it ever come to an end?

Some theoretical physicists, dazzled by the beauty of their equations, have ploughed on, postulating ever-more-exotic notions that have become increasingly difficult to test experimentally, judging them only on their explanatory power and mathematical elegance—important criteria, I agree, but not the traditional benchmarks for validating scientific theories. So, instead of patting ourselves on the back for how far we've come, should we consider the possibility that we might be straying too far from the path of physics?

Many physicists will no doubt argue that these past few years have been tremendously exciting for fundamental physics, considering the widely reported discoveries of the Higgs boson at the Large Hadron Collider in 2012, followed by

gravitational waves at the LIGO (Laser Interferometer Gravitational-Wave Observatory) facilities in the United States in 2016. But the truth is that both these observational discoveries, vital though they are, 'merely' confirm predictions made by theorists a long time ago—fifty years in the case of the Higgs, and a full century for gravitational waves. I know this sounds more than a little dismissive, and I do not want to downplay the extraordinary achievements of the thousands of experimental physicists and engineers who played a part in these two remarkable discoveries. But, when I say 'merely,' I mean that there weren't many physicists who *didn't* expect these experimental confirmations to be made one day. In the case of the Higgs, even though the discovery led to the award of the physics Nobel Prize the following year, it went to the theorists who predicted it back in the 1960s, not to the experimentalists who made the confirming observation.

I guess I should make a more careful distinction here between the discovery of the Higgs boson and detection of gravitational waves. The

former was by no means a foregone conclusion; many physicists, including Stephen Hawking, had doubted its existence before 2012. Gravitational waves, in contrast, were entirely expected, since they had not only been predicted by general relativity, but had been indirectly observed many years ago in the behaviour of binary pulsars (pairs of neutron stars in orbit around each other).

In fact, if I look back over the past three decades and consider some of the exciting breakthroughs and discoveries in fundamental physics, such as the top quark, Bose-Einstein condensates, quantum entanglement, neutron star mergers, and exoplanets, I could argue that none of these was completely unexpected. In fact, only one discovery in physics during this period was truly revolutionary and surprising (to the astronomers who first saw it, if not all cosmologists)—that of dark energy in 1998. Otherwise, when it comes to testing our theories and models at the furthest extremes of fundamental physics—the quantum and the cosmic scales—there has been experimental silence. Many of the ideas and speculative

theories I have discussed in this chapter might well turn out to be correct. But it's worth pointing out that the traditional types of experiments that have served to verify or falsify scientific theories in the past are unlikely to be able to help us in the future to reach a sufficient level of confidence in their veracity.

When the Large Hadron Collider first started running in 2010, it was only the latest in a long line of particle accelerators around the world, going back almost a century, that have been smashing subatomic matter together at increasingly higher energies. Physicists had waited a long time for the LHC and had high hopes that it would help them answer a number of the outstanding questions and remove uncertainties from the Standard Model. But, above all else, it was billed as the accelerator that would find the Higgs boson, and it duly did—surely, a resounding success and a justification for the huge cost of the project. But since then, there has been mounting frustration that nothing more has been discovered—both from scientists in other disciplines envious of the funding that

CERN receives, and from theoretical physicists impatient for confirmation of their latest predictions.

And what of the Higgs discovery itself? What new insights has it given us about the nature of matter? It's worth noting that the Higgs boson is merely the particle manifestation (excitation) of the more fundamental Higgs field—yet another quantum field that pervades all space and an important ingredient of the Standard Model, because the way other particles move through the Higgs field is what gives them their mass. For instance, the W and Z bosons, the carriers of the weak force, would have no mass without it, and would be more like their cousin, the massless photon. But the W and Z do have mass, and it is the Higgs mechanism that explains how they acquire it— through their interaction with the Higgs field in a way that the photon does not. Final proof of the Higgs field's existence was found not by detecting it directly, but indirectly, through the creation of the evanescent quantum of the field: the Higgs boson.

Finding the Higgs was a remarkable achievement. But, in truth, it was a box to be ticked. The Higgs field was bolted onto the Standard Model, which lived to fight another day. The discovery of the Higgs hasn't opened up many new avenues in fundamental physics research, because it didn't progress our understanding beyond what physicists already knew and expected. The Standard Model remains a consistent framework for our understanding of the building blocks of matter, but it is not a fully coherent or predictive unified theory.

Of course, there are still plenty of data to sift through at the LHC from the most recent run, which came to an end in December 2018, so we might still find something new once all the data have been analysed. But the fact remains that there are still a number of outstanding questions to answer, and we may have to look beyond the LHC to do so. These questions include: Why is gravity so much weaker than the other forces? Why are there just three generations of quarks and leptons? And where does the Higgs's own mass come from? Probably most pressing of all,

and therefore most frustrating because it has not been found, is whether we will find any evidence of supersymmetry.

Just because we *want* supersymmetry to be true does not make it so. Sure, it solves many problems and provides useful insights. It is also neat and logical and aesthetically pleasing. But the longer we go without finding any experimental proof of supersymmetry, the more frustrated we become. At the same time, critics of superstring theory grumble that the field continues to be attractive to the brightest minds because it offers jobs. Young researchers feel safer following in the footsteps of their professors and fear that if they don't, they will lose out on funding and career progression. Meanwhile, university physics departments, competing for scarce resources, see research in string theory as a cheap way of working at the forefront of physics. But as long as progress remains slow, with no new experimental evidence emerging to bolster the efforts of those in the field, dissenting voices will grow louder.

Some might argue that if supersymmetry is correct, then we should probably have found evidence of it by now at the LHC. The simplest class of supersymmetry models (what is called constrained minimal supersymmetry) is already looking unlikely. But this does not mean that we give up on it entirely just yet—we may just be looking for it in the wrong place. After all, it is not only on string theorists' wish lists. More 'down-to-earth' particle physicists also want to know if nature is supersymmetric. Supersymmetry allows us to understand a connection between the electroweak force and the strong nuclear force described by QCD. It also links together the matter particles and the force carrier particles. It would even explain why the Higgs boson has the mass that it does. But solving all these problems comes at a price: supersymmetry predicts the existence of a whole host of new particles that have yet to be discovered.

I should of course say that there is a nice added bonus if supersymmetry is true: the lightest of these yet-to-be-observed supersymmetric particles fit the bill as the constituents of dark matter.

REASONS TO BE OPTIMISTIC

Theoretical physicists have not just been sitting around waiting on news from their experimental colleagues. Enthralled by the splendour of their mathematics, they have forged ahead without them. No sooner had the latest version of string theory (M-theory) been proposed in the mid-1990s by Edward Witten than a powerful new idea was developed by Juan Maldacena in 1997. It is known as gauge/gravity duality (or, to give it its technical name, AdS/CFT[8]), and it describes how the strings in string theory relate to the field theories describing the three quantum forces. This mathematical idea has since been developed more generally to tackle problems in other areas of theoretical physics, such as hydrodynamics, quark-gluon plasma, and condensed matter, and Maldacena's paper has become one of the most important pieces of work in modern theoretical physics, having been cited over

8 Which stands for anti–de Sitter/conformal field theory correspondence

17,000 times in other peer-reviewed papers to date.

Powerful ideas like gauge/gravity duality convince many physicists that string theory is the right path to pursue. But even if it turns out not to be the correct theory of quantum gravity, what it has done is provide physicists with a useful and precise mathematical toolbox that has at least shown us there *is* a way of combining quantum mechanics and general relativity consistently, and so it gives us hope that the project of unification is in principle possible. But the fact remains that just because string theory or gauge/gravity duality are mathematically beautiful, this does not make them true.

Where will the final answer come from, then? Maybe from string theory, maybe from the study of black holes, maybe from those working on quantum information theory trying to build quantum computers, or maybe even from condensed matter theory. It is becoming clearer that similar forms of mathematics apply across all these areas. In the search for a correct theory of quantum gravity, we may not even need to quantise

gravity, after all. Maybe trying to *force* quantum field theory and general relativity towards each other is the wrong approach. There is some evidence that quantum field theories may already contain within them the essence of curved spacetime, and that general relativity may be closer to quantum mechanics than we thought.

It would be fascinating to know which of the many ideas and theories in this chapter turn out to be correct and which will be consigned to the rubbish heap of wrong science. For me personally, the biggest unanswered question in physics is one that has vexed me all my professional life: What is the correct interpretation of quantum mechanics? In chapter 5, I touched on a few of the candidate ideas and mentioned that for many physicists this really is a problem for the philosophers to address, since it has not stopped the application of quantum mechanics or slowed the progress of physics. But a growing number of physicists, including yours truly, see the foundations of quantum mechanics as a vitally important field, and suspect that a resolution of the long-standing issue of its interpretation will ultimately

lead to new physics. It may even be linked to one or more of the other outstanding problems in fundamental physics, such as the nature of time or the ultimate theory of quantum gravity.

These problems sometimes seem so difficult to overcome that it wouldn't surprise me if we ended up needing an advanced artificial intelligence to help us. Maybe an AI of our invention will emerge as the next Newton or Einstein, and we may have to accept that our puny human brains are just not smart enough to figure out the ultimate nature of reality on our own.[9]

I have focussed in this chapter on the future of physics, mainly with regard to mathematical physics and physics at the extremes of the very smallest and largest scales. But is this fair? Are these justifiably the true frontiers of physics? Advances in physics are not all about striving to see smaller or further—and the everyday scales, in terms of size and energy, are no less fascinating. In fact, in terms of how physics will transform our lives in the twenty-first century, the real

---

9 Maybe it will inform us that answer is, indeed, 42.

excitement is in areas like condensed matter physics and quantum optics, and in areas in which physics overlaps and merges with chemistry, biology, and engineering. So, rather than delve into some of those topics here, I will use them as examples of how the technological applications of physics are shaping our world and I will explore these, some might say, more 'useful' aspects of the physics enterprise in the next chapter.

CHAPTER 9

THE USEFULNESS OF PHYSICS

Wherever you are right now as you're reading this book, look around you. So much of what we humans have created and built has only been possible thanks to our understanding of the laws of nature: the forces that shape our world and the properties of the matter that these forces act on. It would therefore be impossible to list all of the applications of physics—all the trappings of our modern world that have emerged from the discoveries made by physicists over the centuries[1]—so I will instead focus on two topics. The first is the way physics has underpinned,

---

1 Of course, I am not claiming that this knowledge and understanding has come about exclusively because of the work of physicists, since I could make a similar claim if I were writing about chemistry, or engineering, or mathematics.

overlapped with, and even merged with other disciplines, both pure and applied, and the role it continues to play in a number of exciting new interdisciplinary research fields. The second is a brief look at a selection of new applications that will undoubtedly arise from current physics research, with a particular focus on the exciting prospect of new quantum technologies.

After getting this far in the book you will be forgiven for thinking that physicists' obsession with unifying the mathematical principles that govern the workings of nature is all very well—a testament to humankind's tenacious drive to understand the universe—but so what? Surely, you might think, the discovery of the Higgs boson won't have any sort of direct impact on our daily lives; nor will the hoped-for theory of quantum gravity help to eradicate poverty and disease. But this is not the right way to look at things. Fundamental, curiosity-driven science has, time and again, led to technological advances that have revolutionised our world. Most physics researchers, particularly those working in academia, are not typically motivated by the potential applica-

tions of their work, and if you look back at the great discoveries in science which later proved to have practical benefit, you will find that many of them were made out of a burning desire simply to understand the world and to satisfy scientists' curiosity.

Consider a superficial comparison between physics and engineering. A mechanical or electrical engineering student will study many of the same topics as a physics student, such as Newtonian mechanics, electromagnetism, computing, and the mathematical techniques required for solving certain types of equations that crop up on a regular basis. Indeed, many applied physicists end up working in engineering industries, further blurring the boundaries between the two disciplines. But, typically, a physicist will ask the 'Why?' and the 'How?' questions in order to expose the underlying principles that govern the workings of nature, whereas an engineer is not normally motivated by these deeper principles and will rather put his or her understanding to work, using it to build a better world. Both physicists and engineers are problem solvers,

but they tend to have different motivations for seeking solutions.

Just to offer a particular example, the brilliant engineering success of satellite navigation systems (the US GPS being the most important one over the past few decades) clearly demonstrates the value of pure physics research that underpins the engineering. GPS systems are now such an integral part of our lives that we could not live without them. Not only do we take for granted the fact that we no longer get lost in unfamiliar parts of the world, but GPS has allowed us to see our planet from above and map it with remarkable detail, enabling us to study the way the Earth's climate is changing, or to predict natural phenomena and help with disaster relief. In the future, global positioning satellites will link with AI systems to transform transport, agriculture, and many other industries. And yet without the knowledge that came from fundamental physics research, GPS would not have been possible. For example, the atomic clocks onboard satellites, which are needed to ensure that we can locate them precisely in order to pinpoint our position on the ground, only work

because engineers had to take the quantum nature of atomic vibrations into account, along with the relativistic corrections to the rate of flow of time that Einstein's theories explained.

There are countless other technological examples of physics and engineering overlapping in ways that have changed the world as we know it. And engineers aren't the only people physicists have long had a close working relationship with. Today, many physicists work alongside scientists from a wide range of disciplines, such as medicine, neuroscience, computer science, bioengineering, geology, environmental science, and space science. You will also find physicists applying their logical, numerate, and problem-solving skills outside of science in professions ranging from politics to finance.

WHERE PHYSICS, CHEMISTRY, AND BIOLOGY MEET

Throughout the history of science, there has always been a strong overlap between physics and its sister discipline, chemistry. Indeed, some of

the greatest scientists of all time, most notably Michael Faraday, have been claimed by both subjects. And it isn't just chemistry; the role of physics in biology, in particular, has a fascinating history. The community of physicists interested in biological problems is incredibly diverse, and their work has led to a very vibrant field of research called biophysics. But is biophysics a branch of physics, or is it no more than the application of methods from physics to the problems of biology? Does this distinction even matter? If physics ultimately underpins chemistry and chemical processes, and the phenomena within living organisms are themselves nothing more than complex chemistry, then surely it follows that physics must lie at the heart of biology. After all, everything, living or inanimate, is ultimately made of atoms and is subject to the laws of physics.

In an effort to dig down and identify the fundamental principles that govern the workings of biology, as is their wont, physicists ask, 'What distinguishes life from non-life *made up of the same ingredients*?' The answer is rooted in physics:

life has the ability to maintain itself in a state of low entropy, far from thermal equilibrium, and is able to store and process information. Therefore, one has the feeling that a full understanding of what makes life special has to come from fundamental physics. In writing this, I can imagine my colleagues in chemistry and biology rolling their eyes with exasperation at this typical physicist's sense of self-importance. On the other hand, it is true that many of the early advances in molecular biology and genetics in the twentieth century were made by physicists, such as Leo Szilard, Max Delbrück, and Francis Crick. Crick, in particular, who discovered (with James Watson and Rosalind Franklin) the double helix structure of DNA, was hugely influenced by yet another physicist, Erwin Schrödinger, whose remarkable 1944 book, *What is Life?*, is still relevant today.

On the applied side, physicists have been key to the development of many of the techniques that are used to probe living matter, from X-ray diffraction to MRI scanners. Even the humble microscope, without which no biology lab could

function, was invented by physicists, thanks to hundreds of years of research into the nature of light and the way lenses refract and focus it, culminating in the work of Antonie van Leeuwenhoek and Robert Hooke, who both used the microscope in the seventeenth century to study living organisms. Indeed, if you consider the many contributions to science made by Hooke, you will find that, by today's definition, he is most definitely a physicist rather than a biologist.

One particular new area of research that I have personally become interested in over the past twenty years is called quantum biology. This subject should not be confused with the comments I made earlier about all life being ultimately made of atoms and hence at some basic level subject to the rules of the quantum world like everything else in the universe. That is, of course, taken as a given. Rather, quantum biology refers to recent research in theoretical physics, experimental biology, and biochemistry which suggests that some of the more counterintuitive ideas in quantum mechanics, such as tunneling, superposition, and entanglement, might be playing

an important role inside living cells. Key experimental observations regarding the way enzymes work, or in the process of photosynthesis, seem to require a quantum explanation. This has come as a huge surprise to many scientists who refuse to believe that such delicate and strange behaviour could impact on the machinery of life, and the jury is still out on some of these ideas. But don't forget that life has had almost four billion years of evolution to find any shortcuts that would give it an advantage. If quantum mechanics can make a particular biochemical process or mechanism more efficient, then evolutionary biology will make use of it. It isn't magic, it's just . . . well . . . physics.

THE QUANTUM REVOLUTION CONTINUES

In the twentieth century (and the beginning of the twenty-first), there is no doubt that quantum mechanics has had a profound impact on our lives, despite operating at length scales far smaller than our senses can detect. In describing

the subatomic world so successfully, it underpins not only physics and chemistry, but modern electronics, too. For example, an understanding of the quantum rules that explain how electrons behave in semiconductor materials like silicon has laid the foundation of our technological world. Without an understanding of semiconductors, we would not have developed the transistor and, later, the microchip and the computer. That handheld supercomputer we all carry around with us today (our smartphone), without which many of us would now feel utterly lost, is packed full of electronic magic, none of which would have been possible without quantum mechanics. The same goes for many familiar devices in our homes, from televisions and games consoles to modern LED lighting and smoke detectors, and, of course, the internet. Indeed, the entire telecommunications industry relies on technological applications of quantum mechanics, such as lasers and optical amplifiers. And no modern-day hospital could get by without the applications of quantum mechanics, from MRI, PET, and CT scanners to laser surgery.

And the quantum revolution is only just getting started. We are about to witness many new technological marvels in the coming decades, emerging from current research in quantum physics, such as smart materials and topological materials. Take graphene, for example: single layers of carbon atoms arranged in a hexagonal crystal lattice. Depending on how it is shaped and manipulated, graphene acts as an insulator, a conductor, or even as a semiconductor.

What's more, recent research suggests that two layers of graphene twisted at a particular angle to each other can, under certain conditions of low temperature and an applied weak electric field, behave as a superconductor, through which current can flow with no electrical resistance whatsoever—yet another quantum phenomenon. This technique, known as twistronics, will, it is predicted, have applications in a wide range of electronic devices that have yet to be invented.

And there's so much more. A new generation of devices and technologies are currently being developed that will become ubiquitous within

our lifetimes—devices that can create and manipulate exotic states of matter by utilising the tricks of the quantum world in new ways. Advances in areas such as quantum information theory, quantum optics, and nanotechnology will allow us to develop a range of such devices. For example, highly accurate quantum gravimeters will be able to map tiny changes in the Earth's gravitational field, so that geologists can locate new mineral deposits or locate pipes under roads to minimise disruption when workers need to access them. Quantum cameras will have sensors that let us see *behind* obstacles; quantum imaging will allow non-intrusive mapping of brain activity with the potential to tackle conditions like dementia. Quantum key distribution (QKD) will enable us to exchange information securely from one place to another. Quantum technologies will also help us build artificial molecular machines that can carry out a multitude of tasks.

Medicine in particular is a good example of where the quantum world is likely to have a big impact in the coming years. Down at length

scales even smaller than living cells, we are going to see a range of spectacular new technologies emerging, such as nanoparticles with unique quantum properties that allow them to attach to antibodies to help tackle infections, or to be 'programmed' to replicate only inside tumor cells, and even to take images of cells from the inside. Then, quantum sensors will allow us to make far more precise measurements and help with imaging of individual biomolecules. And with the help of quantum computers, which I will discuss in the next section, we should be able to sequence DNA far more quickly than ever before, as well as solve certain tasks that involve searching through all that 'big data' on every aspect of our health, all the way down to the molecular level.

I am deliberately being very selective in my examples here, as there are thousands of technological and engineering advances in communications, medicine, energy, transport, imaging, and sensing that will come about thanks to physics. But one area does deserve further expansion.

QUANTUM COMPUTERS AND
21ST-CENTURY SCIENCE

If you thought the quantum revolution of the last century was impressive, just wait and see what the rest of the twenty-first century has in store. These advances are not only going to give us smarter toys that some might argue just make our lives more complicated; they will help us address some of the biggest challenges facing humanity and transform our world in as yet unimaginable ways. One of the most exciting future applications of physics, without doubt, is the quantum computer. Such a device would be very different from conventional computers and will be used for a wide range of tasks that are impossible today even with the most powerful supercomputers. Quantum computers are expected to help humanity solve many of the most difficult problems in science, especially if combined with advances in artificial intelligence.

Quantum computers rely in a very direct way on the more counterintuitive features of the quantum world. In classical computing, informa-

tion is stored and processed in the form of 'bits' (which stands for binary digits). A single bit of information can have one of two values: zero or one. Combinations of electronic switches, each one a physical manifestation of a bit of information that is either on or off, are used to make logic gates, the building blocks that make up a logic circuit. In contrast, quantum computers operate on what are called quantum bits, or 'qubits', which are not restricted to holding just one or the other of these binary states. Instead, a qubit can exist in quantum superposition of both zero and one simultaneously, and, as such, can store much more information.

The simplest example of a qubit would be an electron whose quantum spin can point either parallel (call this spin 'up') or antiparallel (spin 'down') to an applied magnetic field. If an additional electromagnetic pulse is then applied, it can flip the electron's spin over from parallel (0) to antiparallel (1). But because an electron is a quantum particle, the electromagnetic pulse can also put it in a superposition of spin up (0) and down (1) at the same time. Two

entangled electrons, can be put into a superposition of four possible quantum states—00, 01, 10 and 11—simultaneously. With many more qubits, complicated quantum logic circuits can be developed.

When multiple qubits are entangled together, they can act coherently and therefore process multiple options simultaneously, which makes them far more powerful and efficient than their classical counterparts. But there are problems with realising such a device. Quantum entangled states are extremely delicate and can only be maintained under special conditions for short periods of time. The challenge is not only to isolate and protect these states from their surrounding environment, which destroys quantum coherence, but to be able to control the input and output of the information that the qubits process. And this gets progressively more difficult the larger the number of entangled qubits. Once the computation has been completed, one of the possible final states in the superposition of the qubits is selected and has to be amplified so as to be readable using a macroscopic (classical)

device, which is just one of the many outstanding problems of implementation that have yet to be solved.

Despite these difficult challenges, many research labs around the world are today in a race to build the first true quantum computer. Not that many years ago, it wasn't even clear if such a device would be possible, but now researchers talk about their dream being realised in the next decade or two, and rudimentary prototypes already exist. There are currently a number of different approaches to building a quantum computer, and it is not yet clear which will be the most practical. Typically, qubits can be created from any subatomic particles that exhibit quantum behaviour and which can be entangled together, such as electrons and photons, or ions suspended in electromagnetic fields, or atoms trapped by laser beams, or special liquids and solids in which the quantum spin of their atomic nuclei can be probed using nuclear magnetic resonance.

Computing giants IBM and Google are currently involved in the race to build the first true

quantum computer, but neither has so far been able to build a stable multi-qubit system that can last long enough to make quantum computing practically viable yet. There are also many smaller start-up companies working on the problem. Some focus on the issue of stability, while others are working on increasing the number of entangled qubits. But progress is being made, and I have no doubt that ubiquitous quantum computing will become a reality in my lifetime.

It is important to point out that it is not just designing the hardware that poses a challenge. Quantum computers will also need their own special software to run on them, and quantum algorithms are still in short supply. The best-known examples are Shor's factorization algorithm and Grover's search algorithm. What has already been demonstrated is that such algorithms would allow quantum computers to outperform their classical counterparts in some surprising ways. They will by no means replace current computers at all tasks but will instead be very well suited to solving particular mathematical problems. We will continue to use the ever-

increasing power and processing speed of classical computers to run our daily lives, especially as we advance on a number of fronts with AI, Cloud technologies, and the Internet of Things (the idea that many devices in our homes and workplaces will be connected and talking to each other). And classical computers will also continue to process the ever-increasing mountains of data we have.

There are problems, however, that even the most powerful classical computers of tomorrow will not be able to solve. The beauty of quantum computers is that their processing speed scales exponentially with the number of qubits. Consider the information content of three non-quantum switches. Each can be either 0 or 1, and so there are eight different combinations: 000, 001, 010, 100, 011, 101, 110, 111. But three entangled qubits allow us to store all eight combinations at once. Each of the three digits is both a 1 and a 0 at the same time. On a classical computer, the amount of information increases exponentially with the number of bits. So N bits means 2^N different states. A quantum computer with N

qubits can make use of all 2^N states at once. The hard part is designing algorithms able to make use of this large information space.

Quantum computers will one day be used to solve problems across a wide range of disciplines, in mathematics, chemistry, medicine, and artificial intelligence. Chemists are eagerly looking forward to the prospect of using quantum computers to model highly complex chemical reactions. In 2016, Google developed a rudimentary quantum device that was able to simulate a hydrogen molecule for the first time, and since then, IBM has succeeded in modelling the behaviour of more-complex molecules. It stands to reason that to understand the nature of the quantum world, you need a quantum simulation. After all, it takes one to know one. Eventually, researchers hope they will be able to use quantum simulations to design synthetic molecules and develop new drugs. In agriculture, chemists could use quantum computers to discover new catalysts for fertiliser that would help reduce greenhouse emissions and improve food production.

In AI, quantum computers will dramatically speed up complex optimisation problems in machine learning. This is vital across a range of industries where increasing productivity and efficiency to maximise output is key. Quantum computers are here likely to revolutionise the field of systems engineering by helping to deliver optimisation insights to streamline output and reduce waste. In the not-so-distant future, quantum engineers will be proficient in a wide range of subjects—from quantum mechanics and electronic engineering to systems engineering, AI, and computer science.

Most exciting of all, for me personally (assuming I am still around to witness it), is that by the mid-twenty-first century, we may well see quantum computers running AI programs that will finally answer some of the most important questions of fundamental physics. They, rather than humans, might be making the big breakthroughs.

There's another reason why I have chosen quantum computing as my example of future technology. A number of theoretical physicists

are pinning their hopes on quantum computing to help them out. This is because a quantum computer by its very nature should be in a position to accurately simulate the quantum world and maybe even help them find the right theory of quantum gravity.

My hope, in the subject matter I have covered in this book, is that I have given you a flavor of what physics has enabled us to understand about our world and how we as a species continue to make use of that knowledge. In the final chapter, I want to zoom out and describe how a physicist, or indeed anyone with scientific training, thinks about the world, and how we come to know what we do about it. In other words, how does this great edifice of science—not just the scientific knowledge itself, but the process of gaining it—work, and why do we trust it?

THINKING LIKE A PHYSICIST

ON HONESTY AND DOUBT

I want to share with you what I think is an interesting story. In 2017, I presented a documentary for BBC television called *Gravity and Me*, in which I explored how our understanding of this fundamental concept that shapes our world has evolved throughout the history of science, from an invisible Newtonian force to the structure of spacetime itself. What made the project even more fun was that we developed a smartphone app that monitors its user's location by recording their GPS coordinates (latitude, longitude, and altitude above sea level) at regular intervals. It then uses this information to calculate the rate at which time is passing for the user. According

to general relativity, time flows at different rates depending on the strength of the gravitational field at that location. Someone on a mountaintop is further away from the centre of the Earth than someone at sea level, and so the mountain dweller feels a very slightly weaker gravitational pull by the Earth. This means that time up on the mountaintop runs ever so slightly faster than at sea level. It's a tiny effect: less than a trillionth of a second faster for each second that passes down at sea level. So, even if someone spent all their life at the top of the mountain, all other factors being equal (impossible, I know), they would live for about a millisecond less than they would have if they'd spent their life at sea level—as measured by a very accurate, but otherwise useless hypothetical clock floating out in space. Compared with the far greater benefits of breathing clean mountain air or a healthy diet and regular exercise, this advantage is somewhat pointless. Still, the physical effect is real, and the app was a bit of fun.

To create the app, we had to take another factor into consideration. As I discussed in chapter 3,

moving clocks run more slowly than stationary ones. So you can slow your time down, relative to someone standing still, by moving. This is an even smaller effect than the one due to gravity, as we do not tend to move about at anywhere near the speed of light, where the effect becomes appreciable. Nevertheless, the app took motion into consideration by checking the user's position at regular intervals, and if he or she had changed location significantly, it could work out how fast they had travelled.

Now, here's the crucial bit. Our planet is not a perfect sphere; it bulges out at the equator. So, someone standing at the equator is further from the centre of the Earth than someone at the North Pole (by about twenty-two kilometres), and so, like the mountain dweller, they should feel a slightly weaker gravitational pull. Therefore, clocks at the Pole, where gravity is stronger, should be ticking ever so slightly slower than those at the equator (this is called general relativistic time dilation). However, the Earth is also spinning, and clocks at the equator are moving faster than clocks at the Pole (as measured

by the adjudicating clock floating out in space)
so equator clocks should tick more slowly than
pole clocks (special relativistic time dilation).
These two effects, due to special and general
relativity, work against each other, so which
one wins? Which clock is ticking more slowly?
I calculated these two effects separately and
found that a clock at the pole ticks, overall, more
slowly because it feels stronger gravity, despite
the clock on the equator moving faster.

All of this cool mathematical information was
incorporated into the app, which implemented
my formulae. An enthusiastic social media cam-
paign meant that we persuaded thousands of
people to download the app and use it before the
programme aired. We even received video dia-
ries from a number of people, such as a pilot and
a mountaineer, who provided a record of their
app's results.

Then we hit a snag.

My very smart producer, Paul Sen, called me
one evening, a week before the editing of the
programme was scheduled to finish and a just
before I was due to record the voice-over—for an

anticipated transmission on the BBC soon after that. He said he'd been reading some material on an online physics forum which suggested that I might have screwed up. I immediately dropped what I was working on and went back to consult my calculations. I also quickly emailed half a dozen colleagues to check what I had done.

I had indeed made a very basic mistake. The two effects—the slowdown of time at the Pole because it feels stronger gravity and the slowdown of time at the equator because it is moving faster—*cancel out exactly*! In fact, all clocks at sea level tick at the same rate, everywhere on Earth, and the time they measure is called International Atomic Time (IAT). The surface of the Earth is a geoid, an equipotential gravitational surface on which the cancellation between the two effects due to special and general relativity is not a coincidence. When our planet first formed billions of years ago, hot and malleable, its spin forced it into a more stable, bulged-out (oblate) shape, thus *ensuring* that all points on its surface are sitting in the same gravitational potential. So time flows at the same rate everywhere, provided it

is measured at sea level—climb higher and your time speeds up, go down below Earth's surface and it slows down.

The numbers that were being produced by my app were wrong, and the formulae had to be corrected. But the problem was more serious than that. I had explained how the app worked in the programme, and my mistake was there for all to see. The documentary couldn't be broadcast in its current form.

I told my producer, who immediately asked the BBC to delay transmission. The easiest solution would of course have been to reshoot the scenes in which I had made the mistake. No one would be any the wiser. But I quickly realised that this gave me a wonderful opportunity to show how real science works. Instead of covering up my blunder, I should come clean about it, to show that in science it's OK to make mistakes. So, we shot some new scenes in which I confessed my error in all its glory and explained why I had been wrong. This admission did not require any particular courage or strength of character on my part, since making mistakes is normally

the way science progresses: they are inevitable, and we learn from them. After all, if we didn't make mistakes, how would we ever discover anything new about the world? This is where science differs from, say, politics. I mean, how often do you hear politicians admitting unequivocally that they are wrong?

The history of science is full of examples of learning from the mistakes of the past, with new hypotheses and theories replacing older ones as our understanding of the workings of nature grows or new empirical evidence becomes available. But how do we explain to wider society the value of this approach: forming a hypothesis, testing it, then rejecting it if it doesn't fit the data? This is all a far cry from so much of public discourse taking place today, particularly on social media, where the loudest voices tend to be from those who value personal opinion and preconceived prejudice over evidence, reproducibility, and rigour.

Is there then a lesson that scientists can teach society, or would we just be accused of arrogance and 'elitism'?

Another trait closely related to our obsession with honesty, and almost unique to scientific research, is the importance of doubt. This trait can sometimes be our own worst enemy when it comes to explaining to wider society how science works. We state that we can never be completely certain about something, that a scientific theory is only our current best guess at an explanation, and that as soon as this theory conflicts with new observations or data, we must be prepared to revise or discard it for something better. But some people then say, 'If you are not sure of anything, how can we trust or believe anything you tell us? Without certainty, what can we cling to?' This response is understandable. It is in our nature to want to *know for sure*, not just to have some temporary 'best guess'.

But to think in this way is to misunderstand how science progresses. The trustworthiness of science comes not from certainty, but from its very openness about its uncertainty, always calling into question what we currently understand and being prepared to replace that knowledge with a deeper understanding if something better

comes along. In other walks of life, this attitude might be regarded as fickle. But not in science. Scientific progress depends on scientists' unwavering commitment to the qualities of honesty and doubt.

Here is another example of the way scientists think that might come as a surprise to you. People are often shocked to hear that many physicists—other than those who had dedicated years of their lives to building the Large Hadron Collider—were hoping that the Higgs boson would *not* be found. You see, not finding the Higgs would have meant that there really was something wrong with the Standard Model, opening the door to exciting new physics. Merely ticking a box to confirm something we already suspect to be true is just not as exciting as finding out that one needs to pursue hitherto unexplored paths of research.

On the other hand, we physicists are sometimes accused by well-meaning but amateur scientists of not being open-minded *enough* to entertain their new theories, such as claims that they have found some flaw in Einstein's relativity. The truth

is: I would love for Einstein to be proven wrong, for that would mean we would need a new and better theory to replace his, just as general relativity improved upon Newtonian gravity. But physicists have been relentlessly checking, prodding, and poking Einstein's ideas for a century now, and still relativity theory keeps coming through with flying colours. A better theory may one day be discovered, of course, explaining everything that relativity does and more. But we have yet to discover it.

And so, as part of our ongoing, centuries-long effort to find ever-more-fundamental explanations of physical phenomena, we keep trying to tear down our existing theories, to test them to destruction. If they survive, then we trust them . . . until something better comes along.

ON THEORY AND KNOWLEDGE

When, in general conversation, someone says they have a theory, what they tend to mean is that they have an opinion on something—a view that may be based on some form of evidence

or observation, but equally well one that is no more than a guess, or a hunch, based on ideology or prejudice or some other belief system. Such a 'theory', which may or may not turn out to be correct, is very different from what we mean by a scientific theory[1]—which also, of course, may or may not be correct, but which, in contrast to a mere opinion, must satisfy a number of important criteria. Firstly, it must put forward an explanation of what we observe, either in nature or in an experiment, and provide evidence for that explanation. Secondly, a scientific theory must be verifiable in accordance with the scientific method: it must be testable, and those tests or observations have to be repeatable. Finally, a good scientific theory makes new predictions about aspects of the world that it explains, which can then also be tested by further observations or experiments.

Our most successful scientific theories, such as relativity, quantum mechanics, the Big Bang

---

1 I am thinking here of theories in the natural sciences rather than in, say, economics or psychology.

theory, Darwinian evolution, plate tectonics, or the germ theory of disease, have all undergone rigorous scrutiny and have all emerged as the best explanations we have. None of these can be dismissed, as one often hears (particularly regarding Darwinian evolution), as 'just a theory'. Such a statement ignores what it means for a scientific theory to be successful—that it has explanatory power, that it is backed up by evidence, and that it makes predictions that can be tested, and yet it remains falsifiable, in the sense that if observations or experimental results contradict its predictions then it cannot be a correct theory, or at best cannot be the whole story.

How then should we counter those who wish to undermine science and the scientific method—those who claim that their opinion should be valued above evidence and that their 'theory' should be given as much credence as the scientific theory it purports to challenge or dispute, without the need to be held to the same standards? While we might find it amusing that some people believe the Earth to be flat, or that the Apollo Moon landings were a hoax, or that the

world was created just a few thousand years ago, what about the people who hold views that not only go against established science, but which are genuinely harmful to society, such as those who deny anthropogenic climate change, those who refuse to vaccinate their children because they believe in a baseless link between MMR and autism, or those who prefer magic and superstition to modern medicine?

I find it frustrating that I do not have a clear answer to these questions. I have devoted half my academic career in physics to my research, trying to understand for myself how the universe works. The other half I have spent teaching, communicating, and explaining what I have learnt. So, I cannot simply absolve myself of any responsibilities to engage and debate scientific issues with the wider public; many of these issues are too important *not* to address. But I also know how hard it is to shift someone's strongly held views on a matter, however misguided I believe they are.

In a very real sense, conspiracy theories are the polar opposite of scientific theories in that

they seek to assimilate whatever evidence there is against them and interpret it in a way that supports rather than repudiates their core idea, thus making them *unfalsifiable*. Many who hold such views will always try to interpret and favour evidence in a way that confirms their pre-existing hypotheses. This is known as confirmation bias. Often, in the case of ideological beliefs, we also hear the term 'cognitive dissonance', whereby someone will feel genuine mental discomfort when confronted with evidence supporting a view contrary to their own. This potent combination of confirmation bias and the avoidance of cognitive dissonance works to reinforce pre-existing beliefs. So, trying to persuade someone in this frame of mind with scientific evidence can often prove to be a waste of time.

Many people, facing an avalanche of widely different views through both the mainstream and social media, understandably find it difficult to know what to believe. How can they tell accurate evidence-based information from fake news? One thing scientists can do is to tackle the issue of false balance. Thus, when almost

every climatologist in the world acknowledges that the Earth's climate is changing rapidly due to humankind's activities and that something needs to be done urgently if we are to prevent its catastrophic consequences, the news media does not need to have a climate change denier provide 'the other side of the argument'. Because when this happens, the public is left with the impression that both points of view are equally valid. Apart from the weight of scientific evidence in their favour, the difference between someone arguing that anthropogenic climate change is real and another denying it is that the former really hopes that he or she is wrong.

A scientist will always admit that *maybe* climate change is not happening; maybe evolution theory is wrong, or relativity is wrong. Maybe gravity won't always pull me down to the ground, and through meditation I should be able to levitate. But these 'maybes' do not mean we don't know. We do know that we will continue to test our theories, and if they hold up we trust them and talk about them with non-scientists. But, as scientists, we are prone to expressing ourselves

in terms of honesty and doubt. Just as the word 'theory' has a different meaning in science compared with everyday conversation, so too does the word 'certain' mean something particular to scientists. Deep down, of course, I *am* actually quite certain that it is impossible to overcome the force of gravity through meditation in order to float off the ground. I am also certain that the Earth is round, that it is billions of years old, and that life evolves.

Am I certain that dark matter exists? Almost.

ON TRUTH

I have often heard it said that there are different ways of getting to 'the truth', or indeed that there are different sorts of truth. No doubt a philosopher or a theologian reading this will regard my simplistic physicist views on the matter as hopelessly naïve, but, for me, absolute truth refers to what is real and what exists independently of human subjectivity. So, when I talk about science as being the quest for truth, I mean that scientists are constantly trying to

get as close as possible to the ultimate nature of things, to an objective reality out there waiting to be discovered and understood. It can sometimes feel as though this objective reality is nothing more than a collection of facts about the world that we discover slowly until we know them all. But remember that in science we can never claim to know something for certain. There is always the chance that, at some later time, we will arrive at an even deeper understanding, taking us closer to that ultimate truth we seek.

In practice, there are many ideas and concepts in science about which we have reached such a level of confidence that we can safely regard them as facts. If I jump off a roof, the Earth will pull me downwards (and I pull it upwards ever-so slightly) according to a simple mathematical relation that is as close as we will ever get to a statement of fact. We do not yet know everything there is to know about gravity, but we do know its effect on objects in our world. If I drop a ball from a height of five metres, I know without having to check with a stopwatch that it

will be in the air for one second before it hits the ground—not two seconds, or half a second, but one second. One day we may find a new theory of quantum gravity, but it will never predict that my ball will take twice or half as long as Newton's equation of motion predicts. That is an absolute truth about the world. There is no philosophical argument, no amount of meditation, no spiritual awakening or religious experience, or gut instinct or political ideology that could ever have told me that a ball dropped from a height of five metres would take one second to hit the ground. But science can tell me.

In a sense, then, the remaining gaps in our understanding of the laws of the universe—the nature of dark matter and dark energy, whether inflation theory is correct, the right interpretation of quantum mechanics, the true nature of time, and so on—are not going to change our understanding of the forces, matter, and energy that make up our everyday world. Future advances in physics are not going to render what we already know obsolete. They will just refine it and give us a deeper understanding.

PHYSICS IS HUMAN

In the end, physicists are like everyone else. We want our ideas and theories to be correct, and we will often defend them in the face of emerging evidence to the contrary. Even the most brilliant physicists have been known to downplay problems with their theories and to amplify their criticisms of a rival idea. Confirmation bias exists in science just as it does in all walks of life, and scientists are not immune to it. We strive for tenure and promotion, to compete for funding, meet project deadlines, 'publish or perish', and work hard to gain the respect of our peers and the approbation of our superiors.

And yet, part of our training in the scientific method is to develop humility and honesty in research to enable us to act against our baser instincts. We learn not to be blinded by our desires or misled by our biases and vested interests. It is sometimes hard to see this if you focus on individuals—and there have been a number of well-documented examples of fraud and corruption in scientific research. But, as research

communities, we have built-in corrective procedures, such as the peer review of scientific papers (and yes, I know this is not an ideal way to evaluate research), and we rigorously train young scientists in the ethical and responsible conduct of research. This means that the scientific method is, by its very nature, self-correcting. It demands repeatability and the continued honest and critical assessment of ideas. Weak theories die out eventually, however hard their advocates try to keep them alive, and even if it sometimes takes a generation or two to free ourselves from the hegemonic shackles of a particularly dominant theory that has passed its 'sell by' date.

The best physicists are often those who have been able to zoom out and free themselves from the prejudices of consensus, fashion, or reputation—even their own. But this is more likely to happen when a theory is already known not to be the final word on a subject, or if there are rival theories, each with its own staunch advocates. And remember that physics, just like all science, is not a democracy. All it takes is one new experimental observation to bring down a

widely accepted theory and replace it with a new one. Thereafter, it is the new theory that must constantly justify itself by being held up against the bright light of observational data.

Many of the more speculative ideas in fundamental physics today—some of which I described in chapter 8—might be considered as failing to meet the requirements of what constitutes a proper scientific theory, because they cannot be checked against experiment. Among these, I would include (for now at least) string theory, loop quantum gravity, black hole entropy, and multiverse theories. And yet, thousands of theoretical physicists around the world are carrying out intensive research in these subjects. Should they stop working in these areas because their theories can't yet be tested? Are they wasting public funds that could be better poured into more 'useful' areas of research? And what drives these physicists on if they have no way of testing their theories? Are they being blinded by the beauty of their equations? It is certainly true that a few physicists have even gone so far to claim that they do not *need* to test their theories against

data, but only against each other, for mathematical consistency and elegance, which strikes me as a dangerous road to go down.

However, being too harsh on these 'searchers in the dark' can also show a lack of imagination and appreciation of the history of ideas in science. When Maxwell wrote down his electric and magnetic field equations and derived from them the wave equation for light, no one knew, not even Maxwell himself, how this knowledge would be used by Heinrich Hertz, Oliver Lodge, Guglielmo Marconi, and others to develop radio. Nor did Einstein anticipate, when developing his theories of relativity, that they would one day be used to give us accurate satellite navigation, which you access using the technological wonders crammed into that supercomputer in your pocket that would have been impossible without the abstract speculations of the early quantum pioneers.

So, the inflationary cosmologists and the string theorists and the loop quantum gravity researchers continue on their quests, and rightly so. Their ideas may turn out to be wrong—or they

may change the course of human history. Or we may have to wait for another Einstein, perhaps even an AI, to help us out of our current confusion. We cannot yet say. But what we *can* say is that, if we ever stop being curious about the universe and investigating how it—and we—came to be, then that is when we stop being human.

The human condition is bountiful beyond measure. We have invented art and poetry and music; we have created religions and political systems; we have built societies, cultures, and empires so rich and complex that no mere mathematical formula could ever encapsulate them. But, if we want to know where we come from, where the atoms in our bodies were formed—the 'why' and 'how' of the world and universe we inhabit—then physics is the path to a true understanding of reality. And with this understanding, we can shape our world and our destiny.

It's not an easy balancing act to try to cover the vast expanse of fundamental physics in a short book such as this, aimed at a lay audience, while still having the space to allow for a level of detail that takes into account the very latest thinking in many topics, and to weave them together. I leave you to be the judge of whether I have succeeded or not. I also wanted to avoid regurgitating some of the overused metaphors and analogies that one so often reads in popular science books. Many of these eventually go out of date, or even turn out to be wrong, given our advancing understanding.

And even if all of this can be achieved, there is another problem.

The island of our physics knowledge is surrounded by an ocean of the yet-to-be-explained, but this island is growing in size all the time. This book is meant to be an exploration of its

shoreline—the very limits of our current understanding. But describing this shoreline concisely and accurately is not an easy undertaking for any one person. Although I can call upon over three decades of research in theoretical physics, a quarter of a century of university teaching and almost as long as a science communicator and author, honing my skills in finding the right language to use to unpack complex concepts, I am nevertheless very aware of my limitations when it comes to fully understanding those areas of physics outside my expertise. I am therefore enormously indebted to colleagues and collaborators for all the many fruitful discussions I have had with them over the years. I am also very grateful to all those who gave up their precious time to read through this manuscript and, in doing so, offered me advice and suggestions that have helped fill in the gaps in my understanding. Often, they suggested subtle changes to the wording to make an explanation more precise while never sacrificing clarity and simplicity.

I have enjoyed being (just a little) polemical here and there in expressing my views on

I have tried as best I can to take into account all the comments and suggestions made by those named above. No doubt there will still be details that some might not entirely agree with, but hopefully there aren't too many of these. One thing I know for sure is that they have made this book far better than I could have hoped for had I not sought their help.

I have for many years now had the pleasure of presenting the BBC Radio 4 series *The Life Scientific*, in which I interview many of the world's leading scientists. This has given me the opportunity to dig a little deeper into the latest ideas in fundamental physics, in particular in esoteric areas such as particle physics and cosmology. As such, I am indebted to Sean Carroll, Frank Close, Paul Davies, Fay Dowker, Carlos Frenk, Peter Higgs, Lawrence Krauss, Roger Penrose, and Carlo Rovelli, all of whom have been wonderful guests on my programme. If there is anything in this book they don't wholeheartedly endorse (and I am sure there will be), then I hope they forgive me. They didn't read the manuscript, but their insights have certainly helped clarify my thinking.

Finally, I owe a huge debt of gratitude to my editor at Princeton University Press, Ingrid Gnerlich, for her enthusiastic support, advice, and suggestions on the book's structure and format that have helped me mould it into its final version, with much appreciated additional help from my copyeditor Annie Gottlieb.

It goes without saying that the biggest 'thank you's go to my lovely wife, Julie, for her patience and support, and to my almost-as-lovely agent, Patrick Walsh—we make a good team.

FURTHER READING

Here is a list of popular science books that expand on the subject matter of this book

GENERAL

Peter Atkins, *Conjuring the Universe: The Origins of the Laws of Nature* (Oxford and New York: Oxford University Press, 2018).

Richard P. Feynman et al., *The Feynman Lectures on Physics*, 3 vols. (Reading, MA: Addison-Wesley, 1963; rev. and ext. ed., 2006; New Millennium ed., New York: Basic Books, 2011); available in full online for free, http://www.feynmanlectures.caltech.edu.

Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford and New York: Oxford University Press, 1989).

Lisa Randall, *Warped Passages: Unraveling the Mysteries of the Universe's Hidden Dimensions*

(London: Allen Lane; New York: HarperCollins, 2005).

Carl Sagan, *The Demon-Haunted World: Science as a Candle in the Dark* (New York: Random House, 1996).

Steven Weinberg, *To Explain the World: The Discovery of Modern Science* (London: Allen Lane; New York, HarperCollins, 2015).

Frank Wilczek, *A Beautiful Question: Finding Nature's Deep Design* (London: Allen Lane; New York: Viking, 2015).

QUANTUM PHYSICS

Jim Al-Khalili, *Quantum: A Guide for the Perplexed* (London: Weidenfeld and Nicolson, 2003).

Philip Ball, *Beyond Weird: Why Everything You Thought You Knew about Quantum Physics Is . . . Different* (London: The Bodley Head; Chicago: University of Chicago Press, 2018).

Adam Becker, *What Is Real? The Unfinished Quest for the Meaning of Quantum Physics* (London: John Murray; New York, Basic Books, 2018).

Sean Carroll, *Something Deeply Hidden: Quantum Worlds and the Emergence of Spacetime* (London: OneWorld; New York: Dutton, 2019).

James T. Cushing, *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony* (Chicago and London: University of Chicago Press, 1994).

David Deutsch, *The Fabric of Reality: Towards a Theory of Everything* (London: Allen Lane; New York: Penguin, 1997).

Richard P. Feynman, *QED: The Strange Theory of Light and Matter* (Princeton and Oxford: Princeton University Press, 1985).

John Gribbin, *Six Impossible Things: The 'Quanta of Solace' and the Mysteries of the Subatomic World* (London: Icon Books, 2019).

Tom Lancaster and Stephen J. Blundell, *Quantum Field Theory for the Gifted Amateur* (Oxford and New York: Oxford University Press, 2014).

David Lindley, *Where Does the Weirdness Go? Why Quantum Mechanics is Strange, but Not as Strange as You Think* (New York: Basic Books, 1996).

N. David Mermin, *Boojums All the Way Through: Communicating Science in a Prosaic Age* (Cambridge, UK, and New York: Cambridge University Press, 1990).

Simon Saunders, Jonathan Barrett, Adrian Kent, and David Wallace, eds., *Many Worlds? Everett, Quantum Theory, & Reality* (Oxford and New York: Oxford University Press, 2010).

PARTICLE PHYSICS

Jim Baggott, *Higgs: The Invention and Discovery of the 'God Particle'* (Oxford and New York: Oxford University Press, 2017).

Jon Butterworth, *A Map of the Invisible: Journeys into Particle Physics* (London: William Heinemann, 2017).

Frank Close, *The New Cosmic Onion: Quarks and the Nature of the Universe* (Boca Raton, FL: CRC Press / Taylor and Francis, 2007).

Gerard 't Hooft, *In Search of the Ultimate Building Blocks* (Cambridge, UK, and New York: Cambridge University Press, 1997).

COSMOLOGY AND RELATIVITY

Sean Carroll, *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself* (New York: Dutton, 2016; London: OneWorld, 2017).

Albert Einstein, *Relativity: The Special and the General Theory,* 100th Anniversary Edition (Princeton, NJ: Princeton University Press, 2015).

Brian Greene, *The Hidden Reality: Parallel Universes and the Deep Laws of the Cosmos* (London; Allen Lane; New York: Alfred A. Knopf, 2011).

Michio Kaku, *Hyperspace: A Scientific Odyssey through Parallel Universes, Time Warps, and the 10th Dimension* (Oxford and New York: Oxford University Press, 1994).

Abraham Pais, *'Subtle is the Lord . . .': The Science and the Life of Albert Einstein* (Oxford and New York: Oxford University Press, 1982).

Christopher Ray, *Time, Space and Philosophy* (London and New York: Routledge, 1991).

Wolfgang Rindler, *Introduction to Special Relativity*, Oxford Science Publications (Oxford and New York: Clarendon Press, 1982).

Edwin F. Taylor and John Archibald Wheeler, *Spacetime Physics* (New York: W. H. Freeman, 1992); free download, http://www.eftaylor.com/spacetimephysics/.

Max Tegmark, *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality* (London: Allen Lane; New York: Alfred A. Knopf, 2014).

Kip S. Thorne, *Black Holes and Time Warps: Einstein's Outrageous Legacy* (New York and London: W. W. Norton, 1994).

THERMODYNAMICS AND INFORMATION

Brian Clegg, *Professor Maxwell's Duplicitous Demon: The Life and Science of James Clerk Maxwell* (London: Icon Books, 2019).

Paul Davies, *The Demon in the Machine: How Hidden Webs of Information Are Finally Solving the Mystery of Life* (London: Allen Lane; New York: Penguin, 2019).

Harvey S. Leff and Andrew F. Rex, eds., *Maxwell's Demon: Entropy, Information, Computing*

(Princeton, NJ: Princeton University Press, 1990).

THE NATURE OF TIME

Julian Barbour, *The End of Time: The Next Revolution in Physics* (Oxford and New York: Oxford University Press, 1999).

Peter Coveney and Roger Highfield, *The Arrow of Time: A Voyage through Science to Solve Time's Greatest Mystery* (London: W. H. Allen; Harper Collins, 1990).

P.C.W. Davies, *The Physics of Time Asymmetry* (Guildford, UK: Surrey University Press; Berkeley, CA: University of California Press, 1974).

James Gleick, *Time Travel: A History* (London: 4th Estate; New York: Pantheon, 2016).

Carlo Rovelli, *The Order of Time*, *trans. Simon Carnell and Erica Segre* (London: Allen Lane; New York: Riverhead, 2018).

Lee Smolin, *Time Reborn: From the Crisis in Physics to the Future of the Universe* (London: Allen Lane; Boston and New York: Houghton Mifflin Harcourt, 2013).

UNIFICATION

Marcus Chown, *The Ascent of Gravity: The Quest to Understand the Force that Explains Everything* (New York: Pegasus, 2017; London: Weidenfeld and Nicolson, 2018).

Frank Close, *The Infinity Puzzle: The Personalities, Politics, and Extraordinary Science behind the Higgs Boson* (Oxford: Oxford University Press; New York: Basic Books, 2011).

Brian Greene, *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory* (London: Jonathan Cape; New York: W. W. Norton, 1999).

Lisa Randall, *Knocking on Heaven's Door: How Physics and Scientific Thinking Illuminate the Universe and the Modern World* (London: Bodley Head; New York: Ecco, 2011).

Carlo Rovelli, *Reality Is Not What It Seems: The Journey to Quantum Gravity*, trans. Simon Carnell and Erica Segre (London: Allen Lane, 2016; New York: Riverhead, 2017).

Lee Smolin, *Three Roads to Quantum Gravity* (London: Weidenfeld and Nicolson, 2000; New York: Basic Books, 2001).

Lee Smolin, *Einstein's Unfinished Revolution: The Search for What Lies Beyond the Quantum* (London: Allen Lane; New York: Penguin, 2019).

Leonard Susskind, *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design* (New York: Little, Brown, 2005).

Frank Wilczek, *The Lightness of Being: Mass, Ether, and the Unification of Forces* (Basic Books, 2008).

INDEX

A NOTE ON THE TYPE

This book has been composed in
Adobe Text and Futura PT

ADOBE TEXT, designed by Robert Slimbach
in 2010 for Adobe Originals, is a Transitional
serif typeface between calligraphic
Renaissance and high-contrast Modern styles.

FUTURA PT, designed by Vladimir Yefimov
and Isabella Chaeva for ParaType (ParaGraph),
is a geometric sans-serif typeface based on
Paul Renner's original design developed by the
Bauer Type Foundry in 1927.

Shining a light on the most profound insights revealed by modern physics, Jim Al-Khalili invites us all to understand what this crucially important science tells us about the universe and the nature of reality itself.

Al-Khalili begins by introducing the fundamental concepts of space, time, energy, and matter, and then describes the three pillars of modern physics—quantum theory, relativity, and thermodynamics—showing how all three must come together if we are ever to have a full understanding of reality. Using wonderful examples and thought-provoking analogies, Al-Khalili illuminates the physics of the extreme cosmic and quantum scales, the speculative frontiers of the field, and the physics that underpins our everyday experiences and technologies, bringing the reader up to speed with the biggest ideas in physics in just a few sittings. Physics is revealed as an intrepid human quest for ever more foundational principles that accurately explain the natural world we see around us, an undertaking guided by core values such as honesty and doubt. The knowledge discovered by physics both empowers and humbles us, and still, physics continues to delve valiantly into the unknown.

Making even the most enigmatic scientific ideas accessible and captivating, this deeply insightful book illuminates why physics matters to everyone and calls one and all to share in the profound adventure of seeking truth in the world around us.

ABOUT THE AUTHOR

Jim Al-Khalili is professor of physics at the University of Surrey. He is one of Britain's best-known science communicators and has written numerous books, including *Quantum: A Guide for the Perplexed*; *The House of Wisdom: How Arabic Science Saved Ancient Knowledge and Gave Us the Renaissance*; and *Life on the Edge: The Coming of Age of Quantum Biology*. He is a fellow of the Royal Society and lives in Southsea, England. TWITTER | @jimalkhalili