1
The roots of science

1.1 The quest for the forces that shape the world

WHAT laws govern our universe? How shall we know them? How may this knowledge help us to comprehend the world and hence guide its actions to our advantage?

Since the dawn of humanity, people have been deeply concerned by questions like these. At first, they had tried to make sense of those influences that do control the world by referring to the kind of understanding that was available from their own lives. They had imagined that whatever or whoever it was that controlled their surroundings would do so as they would themselves strive to control things: originally they had considered their destiny to be under the influence of beings acting very much in accordance with their own various familiar human drives. Such driving forces might be pride, love, ambition, anger, fear, revenge, passion, retribution, loyalty, or artistry. Accordingly, the course of natural events—such as sunshine, rain, storms, famine, illness, or pestilence—was to be understood in terms of the whims of gods or goddesses motivated by such human urges. And the only action perceived as influencing these events would be appeasement of the god-figures.

But gradually patterns of a different kind began to establish their reliability. The precision of the Sun's motion through the sky and its clear relation to the alternation of day with night provided the most obvious example; but also the Sun's positioning in relation to the heavenly orb of stars was seen to be closely associated with the change and relentless regularity of the seasons, and with the attendant clear-cut influence on the weather, and consequently on vegetation and animal behaviour. The motion of the Moon, also, appeared to be tightly controlled, and its phases determined by its geometrical relation to the Sun. At those locations on Earth where open oceans meet land, the tides were noticed to have a regularity closely governed by the position (and phase) of the Moon. Eventually, even the much more complicated apparent motions of the planets began to yield up their secrets, revealing an immense underlying precision and regularity. If the heavens were indeed controlled by the whims of gods, then these gods themselves seemed under the spell of exact mathematical laws.

Likewise, the laws controlling earthly phenomena—such as the daily and yearly changes in temperature, the ebb and flow of the oceans, and the growth of plants—being seen to be influenced by the heavens in this respect at least, shared the mathematical regularity that appeared to guide the gods. But this kind of relationship between heavenly bodies and earthly behaviour would sometimes be exaggerated or misunderstood and would assume an inappropriate importance, leading to the occult and mystical connotations of astrology. It took many centuries before the rigour of scientific understanding enabled the true influences of the heavens to be disentangled from purely suppositional and mystical ones. Yet it had been clear from the earliest times that such influences did indeed exist and that, accordingly, the mathematical laws of the heavens must have relevance also here on Earth.

Seemingly independently of this, there were perceived to be other regularities in the behaviour of earthly objects. One of these was the tendency for all things in one vicinity to move in the same downward direction, according to the influence that we now call *gravity*. Matter was observed to transform, sometimes, from one form into another, such as with the melting of ice or the dissolving of salt, but the total quantity of that matter appeared never to change, which reflects the law that we now refer to as *conservation of mass*. In addition, it was noticed that there are many material bodies with the important property that they retain their shapes, whence the idea of rigid spatial motion arose; and it became possible to understand spatial relationships in terms of a precise, well-defined geometry—the 3-dimensional geometry that we now call *Euclidean*. Moreover, the notion of a 'straight line' in this geometry turned out to be the same as that provided by rays of light (or lines of sight). There was a remarkable precision and beauty to these ideas, which held a considerable fascination for the ancients, just as it does for us today.

Yet, with regard to our everyday lives, the implications of this mathematical precision for the actions of the world often appeared unexciting and limited, despite the fact that the mathematics itself seemed to represent a deep truth. Accordingly, many people in ancient times would allow their imaginations to be carried away by their fascination with the subject and to take them far beyond the scope of what was appropriate. In astrology, for example, geometrical figures also often engendered mystical and occult connotations, such as with the supposed magical powers of pentagrams and heptagrams. And there was an entirely suppositional attempted association between Platonic solids and the basic elementary states of matter (see Fig. 1.1). It would not be for many centuries that the deeper understanding that we presently have, concerning the actual

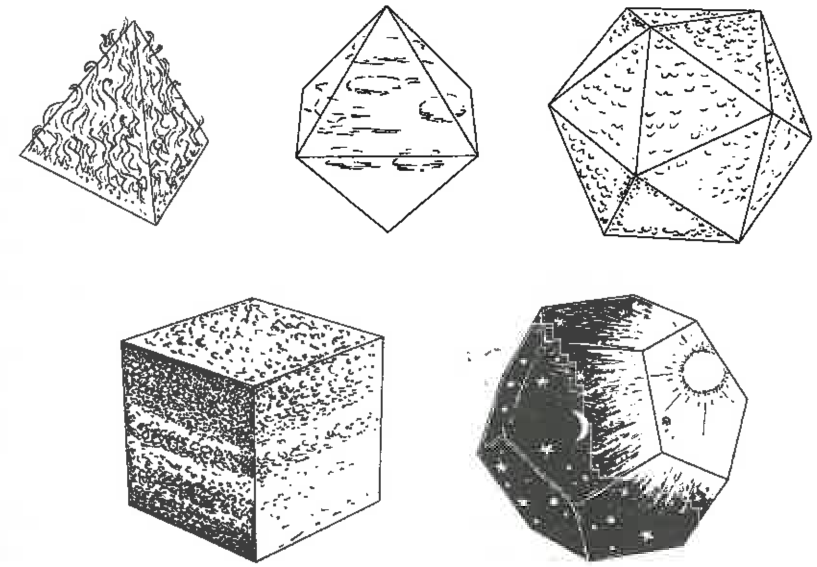

Fig. 1.1 A fanciful association, made by the ancient Greeks, between the five Platonic solids and the four 'elements' (fire, air, water, and earth), together with the heavenly firmament represented by the dodecahedron.

relationships between mass, gravity, geometry, planetary motion, and the behaviour of light, could come about.

1.2 Mathematical truth

The first steps towards an understanding of the real influences controlling Nature required a disentangling of the true from the purely suppositional. But the ancients needed to achieve something else first, before they would be in any position to do this reliably for their understanding of Nature. What they had to do first was to discover how to disentangle the true from the suppositional in *mathematics*. A procedure was required for telling whether a given mathematical assertion is or is not to be trusted as true. Until that preliminary issue could be settled in a reasonable way, there would be little hope of seriously addressing those more difficult problems concerning forces that control the behaviour of the world and whatever their relations might be to mathematical truth. This realization that the key to the understanding of Nature lay within an unassailable mathematics was perhaps the first major breakthrough in science.

Although mathematical truths of various kinds had been surmised since ancient Egyptian and Babylonian times, it was not until the great Greek philosophers Thales of Miletus (*c.*625–547 BC) and

Pythagoras[1*] of Samos (*c*.572–497 BC) began to introduce the notion of *mathematical proof* that the first firm foundation stone of mathematical understanding—and therefore of science itself—was laid. Thales may have been the first to introduce this notion of proof, but it seems to have been the Pythagoreans who first made important use of it to establish things that were not otherwise obvious. Pythagoras also appeared to have a strong vision of the importance of *number*, and of arithmetical concepts, in governing the actions of the physical world. It is said that a big factor in this realization was his noticing that the most beautiful harmonies produced by lyres or flutes corresponded to the simplest fractional ratios between the lengths of vibrating strings or pipes. He is said to have introduced the 'Pythagorean scale', the numerical ratios of what we now know to be frequencies determining the principal intervals on which Western music is essentially based.[2] The famous *Pythagorean theorem*, asserting that the square on the hypotenuse of a right-angled triangle is equal to the sum of the squares on the other two sides, perhaps more than anything else, showed that indeed there is a precise relationship between the arithmetic of numbers and the geometry of physical space (see Chapter 2).

He had a considerable band of followers—the *Pythagoreans*—situated in the city of Croton, in what is now southern Italy, but their influence on the outside world was hindered by the fact that the members of the Pythagorean brotherhood were all sworn to secrecy. Accordingly, almost all of their detailed conclusions have been lost. Nonetheless, some of these conclusions were leaked out, with unfortunate consequences for the 'moles'—on at least one occasion, death by drowning!

In the long run, the influence of the Pythagoreans on the progress of human thought has been enormous. For the first time, with mathematical proof, it was possible to make significant assertions of an unassailable nature, so that they would hold just as true even today as at the time that they were made, no matter how our knowledge of the world has progressed since then. The truly timeless nature of mathematics was beginning to be revealed.

But what is a mathematical proof? A proof, in mathematics, is an impeccable argument, using only the methods of pure logical reasoning, which enables one to infer the validity of a given mathematical assertion from the pre-established validity of other mathematical assertions, or from some particular primitive assertions—the *axioms*—whose validity is taken to be self-evident. Once such a mathematical assertion has been established in this way, it is referred to as a *theorem*.

Many of the theorems that the Pythagoreans were concerned with were geometrical in nature; others were assertions simply about numbers. Those

*Notes, indicated in the text by superscript numbers, are gathered at the ends of the chapter (in this case on p. 23).

that were concerned merely with numbers have a perfectly unambiguous validity today, just as they did in the time of Pythagoras. What about the *geometrical* theorems that the Pythagoreans had obtained using their procedures of mathematical proof? They too have a clear validity today, but now there is a complicating issue. It is an issue whose nature is more obvious to us from our modern vantage point than it was at that time of Pythagoras. The ancients knew of only one kind of geometry, namely that which we now refer to as *Euclidean geometry*, but now we know of many other types. Thus, in considering the geometrical theorems of ancient Greek times, it becomes important to specify that the notion of geometry being referred to is indeed Euclid's geometry. (I shall be more explicit about these issues in §2.4, where an important example of non-Euclidean geometry will be given.)

Euclidean geometry is a specific mathematical structure, with its own specific axioms (including some less assured assertions referred to as postulates), which provided an excellent approximation to a particular aspect of the physical world. That was the aspect of reality, well familiar to the ancient Greeks, which referred to the laws governing the geometry of rigid objects and their relations to other rigid objects, as they are moved around in 3-dimensional space. Certain of these properties were so familiar and self-consistent that they tended to become regarded as 'self-evident' mathematical truths and were taken as axioms (or postulates). As we shall be seeing in Chapters 17–19 and §§27.8,11, Einstein's general relativity—and even the Minkowskian spacetime of special relativity—provide geometries for the physical universe that are different from, and yet more accurate than, the geometry of Euclid, despite the fact that the Euclidean geometry of the ancients was already extraordinarily accurate. Thus, we must be careful, when considering geometrical assertions, whether to trust the 'axioms' as being, in any sense, actually *true*.

But what does 'true' mean, in this context? The difficulty was well appreciated by the great ancient Greek philosopher Plato, who lived in Athens from *c*.429 to 347 BC, about a century and a half after Pythagoras. Plato made it clear that the mathematical propositions—the things that could be regarded as unassailably true—referred not to actual physical objects (like the approximate squares, triangles, circles, spheres, and cubes that might be constructed from marks in the sand, or from wood or stone) but to certain idealized entities. He envisaged that these ideal entities inhabited a different world, distinct from the physical world. Today, we might refer to this world as the *Platonic world of mathematical forms*. Physical structures, such as squares, circles, or triangles cut from papyrus, or marked on a flat surface, or perhaps cubes, tetrahedra, or spheres carved from marble, might conform to these ideals very closely, but only approximately. The actual *mathematical* squares, cubes, circles, spheres, triangles, etc., would not be part of the physical world, but would be inhabitants of Plato's idealized mathematical world of forms.

1.3 Is Plato's mathematical world 'real'?

This was an extraordinary idea for its time, and it has turned out to be a very powerful one. But does the Platonic mathematical world actually exist, in any meaningful sense? Many people, including philosophers, might regard such a 'world' as a complete fiction—a product merely of our unrestrained imaginations. Yet the Platonic viewpoint is indeed an immensely valuable one. It tells us to be careful to distinguish the precise mathematical entities from the approximations that we see around us in the world of physical things. Moreover, it provides us with the blueprint according to which modern science has proceeded ever since. Scientists will put forward models of the world—or, rather, of certain aspects of the world—and these models may be tested against previous observation and against the results of carefully designed experiment. The models are deemed to be appropriate if they survive such rigorous examination and if, in addition, they are internally consistent structures. The important point about these models, for our present discussion, is that they are basically purely abstract *mathematical* models. The very question of the internal consistency of a scientific model, in particular, is one that requires that the model be precisely specified. The required precision demands that the model be a mathematical one, for otherwise one cannot be sure that these questions have well-defined answers.

If the model itself is to be assigned any kind of 'existence', then this existence is located within the Platonic world of mathematical forms. Of course, one might take a contrary viewpoint: namely that the model is itself to have existence only within our various *minds*, rather than to take Plato's world to be in any sense absolute and 'real'. Yet, there is something important to be gained in regarding mathematical structures as having a reality of their own. For our individual minds are notoriously imprecise, unreliable, and inconsistent in their judgements. The precision, reliability, and consistency that are required by our scientific theories demand something beyond any one of our individual (untrustworthy) minds. In mathematics, we find a far greater robustness than can be located in any particular mind. Does this not point to something outside ourselves, with a reality that lies beyond what each individual can achieve?

Nevertheless, one might still take the alternative view that the mathematical world has no independent existence, and consists merely of certain ideas which have been distilled from our various minds and which have been found to be totally trustworthy and are agreed by all.

Yet even this viewpoint seems to leave us far short of what is required. Do we mean 'agreed by all', for example, or 'agreed by those who are in their right minds', or 'agreed by all those who have a Ph.D. in mathematics' (not much use in Plato's day) and who have a right to venture an 'authoritative' opinion? There seems to be a danger of circularity here; for to judge whether or not someone is 'in his or her right mind' requires some external standard. So also does the meaning of 'authoritative', unless some standard of an unscientific nature such as 'majority opinion' were to be adopted (and it should be made clear that majority opinion, no matter how important it may be for democratic government, should in no way be used as the criterion for scientific acceptability). Mathematics itself indeed seems to have a robustness that goes far beyond what any individual mathematician is capable of perceiving. Those who work in this subject, whether they are actively engaged in mathematical research or just using results that have been obtained by others, usually feel that they are merely explorers in a world that lies far beyond themselves—a world which possesses an objectivity that transcends mere opinion, be that opinion their own or the surmise of others, no matter how expert those others might be.

It may be helpful if I put the case for the actual existence of the Platonic world in a different form. What I mean by this 'existence' is really just the objectivity of mathematical truth. Platonic existence, as I see it, refers to the existence of an objective external standard that is not dependent upon our individual opinions nor upon our particular culture. Such 'existence' could also refer to things other than mathematics, such as to morality or aesthetics (cf. §1.5), but I am here concerned just with mathematical objectivity, which seems to be a much clearer issue.

Let me illustrate this issue by considering one famous example of a mathematical truth, and relate it to the question of 'objectivity'. In 1637, Pierre de Fermat made his famous assertion now known as 'Fermat's Last Theorem' (that no positive *n*th power[3] of an integer, i.e. of a whole number, can be the sum of two other positive *n*th powers if n is an integer greater than 2), which he wrote down in the margin of his copy of the *Arithmetica*, a book written by the 3rd-century Greek mathematician Diophantos. In this margin, Fermat also noted: 'I have discovered a truly marvellous proof of this, which this margin is too narrow to contain.' Fermat's mathematical assertion remained unconfirmed for over 350 years, despite concerted efforts by numerous outstanding mathematicians. A proof was finally published in 1995 by Andrew Wiles (depending on the earlier work of various other mathematicians), and this proof has now been accepted as a valid argument by the mathematical community.

Now, do we take the view that Fermat's assertion was always true, long before Fermat actually made it, or is its validity a purely cultural matter, dependent upon whatever might be the subjective standards of the community of human mathematicians? Let us try to suppose that the validity of the Fermat assertion is in fact a subjective matter. Then it would not be an absurdity for some other mathematician X to have come up with an actual and specific counter-example to the Fermat assertion, so long as X had done this before the date of 1995.[4] In such a circumstance, the mathematical community would have to accept the correctness of X's counter-example. From then on, any effort on the part of Wiles to prove the Fermat assertion would have to be fruitless, for the reason that X had got his argument in first and, as a result, the Fermat assertion would now be false! Moreover, we could ask the further question as to whether, consequent upon the correctness of X's forthcoming counter-example, Fermat himself would necessarily have been mistaken in believing in the soundness of his 'truly marvellous proof', at the time that he wrote his marginal note. On the subjective view of mathematical truth, it could possibly have been the case that Fermat had a valid proof (which would have been accepted as such by his peers at the time, had he revealed it) and that it was Fermat's secretiveness that allowed the possibility of X later obtaining a counter-example! I think that virtually all mathematicians, irrespective of their professed attitudes to 'Platonism', would regard such possibilities as patently absurd.

Of course, it might still be the case that Wiles's argument in fact contains an error and that the Fermat assertion is indeed false. Or there could be a fundamental error in Wiles's argument but the Fermat assertion is true nevertheless. Or it might be that Wiles's argument is correct in its essentials while containing 'non-rigorous steps' that would not be up to the standard of some future rules of mathematical acceptability. But these issues do not address the point that I am getting at here. The issue is the objectivity of the Fermat assertion itself, not whether anyone's particular demonstration of it (or of its negation) might happen to be convincing to the mathematical community of any particular time.

It should perhaps be mentioned that, from the point of view of mathematical logic, the Fermat assertion is actually a mathematical statement of a particularly simple kind,[5] whose objectivity is especially apparent. Only a tiny minority[6] of mathematicians would regard the truth of such assertions as being in any way 'subjective'—although there might be some subjectivity about the types of argument that would be regarded as being convincing. However, there are other kinds of mathematical assertion whose truth could plausibly be regarded as being a 'matter of opinion'. Perhaps the best known of such assertions is the *axiom of choice*. It is not important for us, now, to know what the axiom of choice is. (I shall describe it in §16.3.) It is cited here only as an example. Most mathematicians would probably regard the axiom of choice as 'obviously true', while

others may regard it as a somewhat questionable assertion which might even be false (and I am myself inclined, to some extent, towards this second viewpoint). Still others would take it as an assertion whose 'truth' is a mere matter of opinion or, rather, as something which can be taken one way or the other, depending upon which system of axioms and rules of procedure (a 'formal system'; see §16.6) one chooses to adhere to. Mathematicians who support this final viewpoint (but who accept the objectivity of the truth of particularly clear-cut mathematical statements, like the Fermat assertion discussed above) would be relatively weak Platonists. Those who adhere to objectivity with regard to the truth of the axiom of choice would be stronger Platonists.

I shall come back to the axiom of choice in §16.3, since it has some relevance to the mathematics underlying the behaviour of the physical world, despite the fact that it is not addressed much in physical theory. For the moment, it will be appropriate not to worry overly about this issue. If the axiom of choice can be settled one way or the other by some appropriate form of unassailable mathematical reasoning,[7] then its truth is indeed an entirely objective matter, and either it belongs to the Platonic world or its negation does, in the sense that I am interpreting this term 'Platonic world'. If the axiom of choice is, on the other hand, a mere matter of opinion or of arbitrary decision, then the Platonic world of absolute mathematical forms contains neither the axiom of choice nor its negation (although it could contain assertions of the form 'such-and-such follows from the axiom of choice' or 'the axiom of choice is a theorem according to the rules of such-and-such mathematical system').

The mathematical assertions that can belong to Plato's world are precisely those that are objectively true. Indeed, I would regard mathematical objectivity as really what mathematical Platonism is all about. To say that some mathematical assertion has a Platonic existence is merely to say that it is true in an objective sense. A similar comment applies to mathematical *notions*—such as the concept of the number 7, for example, or the rule of multiplication of integers, or the idea that some set contains infinitely many elements—all of which have a Platonic existence because they are objective notions. To my way of thinking, Platonic existence is simply a matter of objectivity and, accordingly, should certainly not be viewed as something 'mystical' or 'unscientific', despite the fact that some people regard it that way.

As with the axiom of choice, however, questions as to whether some particular proposal for a mathematical entity is or is not to be regarded as having objective existence can be delicate and sometimes technical. Despite this, we certainly need not be mathematicians to appreciate the general robustness of many mathematical concepts. In Fig. 1.2, I have depicted various small portions of that famous mathematical entity known

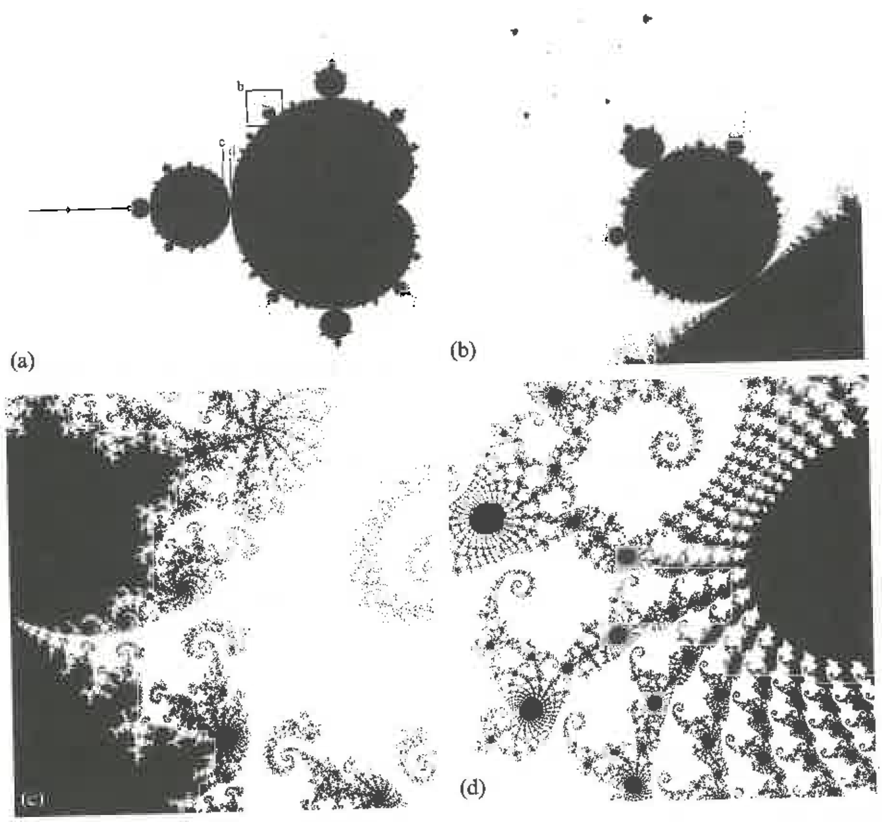

Fig. 1.2 (a) The Mandelbrot set. (b), (c), and (d) Some details, illustrating blow-ups of those regions correspondingly marked in Fig. 1.2a, magnified by respective linear factors 11.6, 168.9, and 1042 (and caps 300, 300, 200, 200; see Note 4.1).

as the *Mandelbrot set*. The set has an extraordinarily elaborate structure, but it is not of any human design. Remarkably, this structure is defined by a mathematical rule of particular simplicity. We shall come to this explicitly in §4.5, but it would distract us from our present purposes if I were to try to provide this rule in detail now.

The point that I wish to make is that no one, not even Benoit Mandelbrot himself when he first caught sight of the incredible complications in the fine details of the set, had any real preconception of the set's extraordinary richness. The Mandelbrot set was certainly no invention of any human mind. The set is just objectively there in the mathematics itself. If it has meaning to assign an actual existence to the Mandelbrot set, then that existence is not within our minds, for no one can fully comprehend the set's

endless variety and unlimited complication. Nor can its existence lie within the multitude of computer printouts that begin to capture some of its incredible sophistication and detail, for at best those printouts capture but a shadow of an approximation to the set itself. Yet it has a robustness that is beyond any doubt; for the same structure is revealed—in all its perceivable details, to greater and greater fineness the more closely it is examined—independently of the mathematician or computer that examines it. Its existence can only be within the Platonic world of mathematical forms.

I am aware that there will still be many readers who find difficulty with assigning any kind of actual existence to mathematical structures. Let me make the request of such readers that they merely broaden their notion of what the term 'existence' can mean to them. The mathematical forms of Plato's world clearly do not have the same kind of existence as do ordinary physical objects such as tables and chairs. They do not have spatial locations; nor do they exist in time. Objective mathematical notions must be thought of as timeless entities and are not to be regarded as being conjured into existence at the moment that they are first humanly perceived. The particular swirls of the Mandelbrot set that are depicted in Fig. 1.2c or 1.2d did not attain their existence at the moment that they were first seen on a computer screen or printout. Nor did they come about when the general idea behind the Mandelbrot set was first humanly put forth—not actually first by Mandelbrot, as it happened, but by R. Brooks and J. P. Matelski, in 1981, or perhaps earlier. For certainly neither Brooks nor Matelski, nor initially even Mandelbrot himself, had any real conception of the elaborate detailed designs that we see in Fig. 1.2c and 1.2d. Those designs were already 'in existence' since the beginning of time, in the potential timeless sense that they would necessarily be revealed precisely in the form that we perceive them today, no matter at what time or in what location some perceiving being might have chosen to examine them.

1.4 Three worlds and three deep mysteries

Thus, mathematical existence is different not only from physical existence but also from an existence that is assigned by our mental perceptions. Yet there is a deep and mysterious connection with each of those other two forms of existence: the physical and the mental. In Fig. 1.3, I have schematically indicated all of these three forms of existence—the physical, the mental, and the Platonic mathematical—as entities belonging to three separate 'worlds', drawn schematically as spheres. The mysterious connections between the worlds are also indicated, where in drawing the diagram

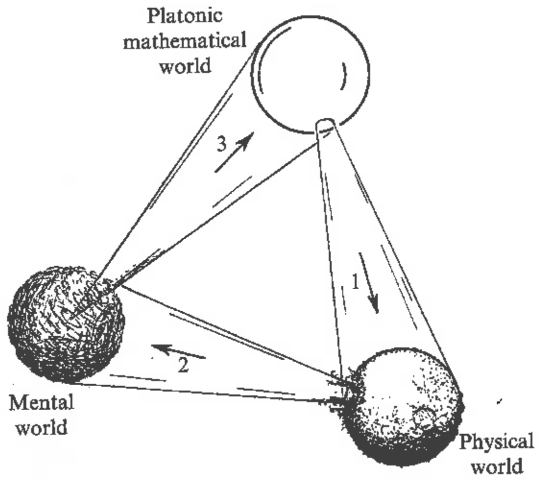

Fig. 1.3 Three 'worlds'—the Platonic mathematical, the physical, and the mental—and the three profound mysteries in the connections between them.

I have imposed upon the reader some of my beliefs, or prejudices, concerning these mysteries.

It may be noted, with regard to the *first* of these mysteries—relating the Platonic mathematical world to the physical world—that I am allowing that only a small part of the world of mathematics need have relevance to the workings of the physical world. It is certainly the case that the vast preponderance of the activities of pure mathematicians today has no obvious connection with physics, nor with any other science (cf. §34.9), although we may be frequently surprised by unexpected important applications. Likewise, in relation to the *second* mystery, whereby mentality comes about in association with certain physical structures (most specifically, healthy, wakeful human brains), I am not insisting that the majority of physical structures need induce mentality. While the brain of a cat may indeed evoke mental qualities, I am not requiring the same for a rock. Finally, for the *third* mystery, I regard it as self-evident that only a small fraction of our mental activity need be concerned with absolute mathematical truth! (More likely we are concerned with the multifarious irritations, pleasures, worries, excitements, and the like, that fill our daily lives.) These three facts are represented in the smallness of the base of the connection of each world with the next, the worlds being taken in a clockwise sense in the diagram. However, it is in the encompassing of each entire world within the scope of its connection with the world preceding it that I am revealing my prejudices.

Thus, according to Fig. 1.3, the entire physical world is depicted as being governed according to mathematical laws. We shall be seeing in later chapters that there is powerful (but incomplete) evidence in support of this contention. On this view, everything in the physical universe is indeed

governed in completely precise detail by mathematical principles—perhaps by equations, such as those we shall be learning about in chapters to follow, or perhaps by some future mathematical notions fundamentally different from those which we would today label by the term 'equations'. If this is right, then even our own physical actions would be entirely subject to such ultimate mathematical control, where 'control' might still allow for some random behaviour governed by strict probabilistic principles.

Many people feel uncomfortable with contentions of this kind, and I must confess to having some unease with it myself. Nonetheless, my personal prejudices are indeed to favour a viewpoint of this general nature, since it is hard to see how any line can be drawn to separate physical actions under mathematical control from those which might lie beyond it. In my own view, the unease that many readers may share with me on this issue partly arises from a very limited notion of what 'mathematical control' might entail. Part of the purpose of this book is to touch upon, and to reveal to the reader, some of the extraordinary richness, power, and beauty that can spring forth once the right mathematical notions are hit upon.

In the Mandelbrot set alone, as illustrated in Fig. 1.2, we can begin to catch a glimpse of the scope and beauty inherent in such things. But even these structures inhabit a very limited corner of mathematics as a whole, where behaviour is governed by strict computational control. Beyond this corner is an incredible potential richness. How do I really feel about the possibility that all my actions, and those of my friends, are ultimately governed by mathematical principles of this kind? I can live with that. I would, indeed, prefer to have these actions controlled by something residing in some such aspect of Plato's fabulous mathematical world than to have them be subject to the kind of simplistic base motives, such as pleasure-seeking, personal greed, or aggressive violence, that many would argue to be the implications of a strictly scientific standpoint.

Yet, I can well imagine that a good many readers will still have difficulty in accepting that all actions in the universe could be entirely subject to mathematical laws. Likewise, many might object to two other prejudices of mine that are implicit in Fig. 1.3. They might feel, for example, that I am taking too hard-boiled a scientific attitude by drawing my diagram in a way that implies that all of mentality has its roots in physicality. This is indeed a prejudice, for while it is true that we have no reasonable scientific evidence for the existence of 'minds' that do not have a physical basis, we cannot be completely sure. Moreover, many of a religious persuasion would argue strongly for the possibility of physically independent minds and might appeal to what they regard as powerful evidence of a different kind from that which is revealed by ordinary science.

A further prejudice of mine is reflected in the fact that in Fig. 1.3 I have represented the entire Platonic world to be within the compass of mentality. This is intended to indicate that—at least in principle—there are no mathematical truths that are beyond the scope of reason. Of course, there are mathematical statements (even straightforward arithmetical addition sums) that are so vastly complicated that no one could have the mental fortitude to carry out the necessary reasoning. However, such things would be *potentially* within the scope of (human) mentality and would be consistent with the meaning of Fig. 1.3 as I have intended to represent it. One must, nevertheless, consider that there might be other mathematical statements that lie outside even the potential compass of reason, and these would violate the intention behind Fig. 1.3. (This matter will be considered at greater length in §16.6, where its relation to Gödel's famous incompleteness theorem will be discussed.)[8]

In Fig. 1.4, as a concession to those who do not share all my personal prejudices on these matters, I have redrawn the connections between the three worlds in order to allow for all three of these possible violations of my prejudices. Accordingly, the possibility of physical action beyond the scope of mathematical control is now taken into account. The diagram also allows for the belief that there might be mentality that is not rooted in physical structures. Finally, it permits the existence of true mathematical assertions whose truth is in principle inaccessible to reason and insight.

This extended picture presents further potential mysteries that lie even beyond those which I have allowed for in my own preferred picture of the world, as depicted in Fig. 1.3. In my opinion, the more tightly organized scientific viewpoint of Fig. 1.3 has mysteries enough. These mysteries are not removed by passing to the more relaxed scheme of Fig. 1.4. For it

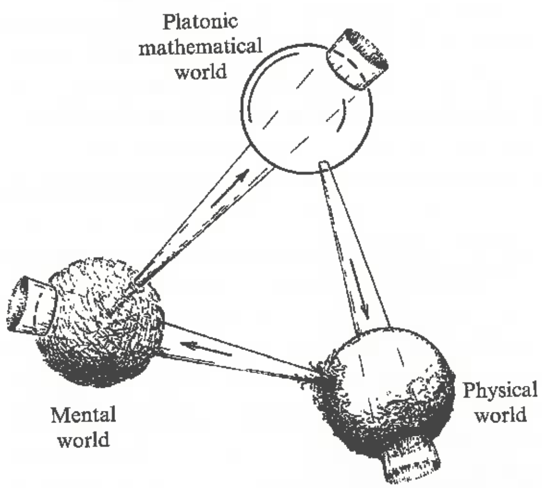

Fig. 1.4 A redrawing of Fig. 1.3 in which violations of three of the prejudices of the author are allowed for.

remains a deep puzzle why mathematical laws should apply to the world with such phenomenal precision. (We shall be glimpsing something of the extraordinary accuracy of the basic physical theories in §19.8, §26.7, and §27.13.) Moreover, it is not just the precision but also the subtle sophistication and mathematical beauty of these successful theories that is profoundly mysterious. There is also an undoubted deep mystery in how it can come to pass that appropriately organized physical material—and here I refer specifically to living human (or animal) brains—can somehow conjure up the mental quality of conscious awareness. Finally, there is also a mystery about how it is that we perceive mathematical truth. It is not just that our brains are programmed to 'calculate' in reliable ways. There is something much more profound than that in the insights that even the humblest among us possess when we appreciate, for example, the actual meanings of the terms 'zero', 'one', 'two', 'three', 'four', etc.[9]

Some of the issues that arise in connection with this third mystery will be our concern in the next chapter (and more explicitly in §§16.5,6) in relation to the notion of *mathematical proof*. But the main thrust of this book has to do with the first of these mysteries: the remarkable relationship between mathematics and the actual behaviour of the physical world. No proper appreciation of the extraordinary power of modern science can be achieved without at least some acquaintance with these mathematical ideas. No doubt, many readers may find themselves daunted by the prospect of having to come to terms with such mathematics in order to arrive at this appreciation. Yet, I have the optimistic belief that they may not find all these things to be so bad as they fear. Moreover, I hope that I may persuade many readers that, despite what she or he may have previously perceived, mathematics can be fun!

I shall not be especially concerned here with the second of the mysteries depicted in Figs. 1.3 and 1.4, namely the issue of how it is that mentality—most particularly conscious awareness—can come about in association with appropriate physical structures (although I shall touch upon this deep question in §34.7). There will be enough to keep us busy in exploring the physical universe and its associated mathematical laws. In addition, the issues concerning mentality are profoundly contentious, and it would distract from the purpose of this book if we were to get embroiled in them. Perhaps one comment will not be amiss here, however. This is that, in my own opinion, there is little chance that any deep understanding of the nature of the mind can come about without our first learning much more about the very basis of physical reality. As will become clear from the discussions that will be presented in later chapters, I believe that major revolutions are required in our physical understanding. Until these revolutions have come to pass, it is, in my view, greatly optimistic to expect that much real progress can be made in understanding the actual nature of mental processes.[10]

1.5 The Good, the True, and the Beautiful

In relation to this, there is a further set of issues raised by Figs. 1.3 and 1.4. I have taken Plato's notion of a 'world of ideal forms' only in the limited sense of mathematical forms. Mathematics is crucially concerned with the particular ideal of *Truth*. Plato himself would have insisted that there are two other fundamental absolute ideals, namely that of the *Beautiful* and of the *Good*. I am not at all averse to admitting to the existence of such ideals, and to allowing the Platonic world to be extended so as to contain absolutes of this nature.

Indeed, we shall later be encountering some of the remarkable interrelations between truth and beauty that both illuminate and confuse the issues of the discovery and acceptance of physical theories (see §§34.2,5,9 particularly; see also Fig. 34.1). Moreover, quite apart from the undoubted (though often ambiguous) role of beauty for the mathematics underlying the workings of the physical world, aesthetic criteria are fundamental to the development of mathematical ideas for their own sake, providing both the drive towards discovery and a powerful guide to truth. I would even surmise that an important element in the mathematician's common conviction that an external Platonic world actually has an existence independent of ourselves comes from the extraordinary unexpected hidden beauty that the ideas themselves so frequently reveal.

Of less obvious relevance here—but of clear importance in the broader context—is the question of an absolute ideal of morality: what is good and what is bad, and how do our minds perceive these values? Morality has a profound connection with the mental world, since it is so intimately related to the values assigned by conscious beings and, more importantly, to the very presence of consciousness itself. It is hard to see what morality might mean in the absence of sentient beings. As science and technology progress, an understanding of the physical circumstances under which mentality is manifested becomes more and more relevant. I believe that it is more important than ever, in today's technological culture, that scientific questions should not be divorced from their moral implications. But these issues would take us too far afield from the immediate scope of this book. We need to address the question of separating true from false before we can adequately attempt to apply such understanding to separate good from bad.

There is, finally, a further mystery concerning Fig. 1.3, which I have left to the last. I have deliberately drawn the figure so as to illustrate a paradox. How can it be that, in accordance with my own prejudices, each world appears to encompass the next one in its entirety? I do not regard this issue as a reason for abandoning my prejudices, but merely for demonstrating the presence of an even deeper mystery that transcends those which I have been pointing to above. There may be a sense in

which the three worlds are not separate at all, but merely reflect, individually, aspects of a deeper truth about the world as a whole of which we have little conception at the present time. We have a long way to go before such matters can be properly illuminated.

I have allowed myself to stray too much from the issues that will concern us here. The main purpose of this chapter has been to emphasize the central importance that mathematics has in science, both ancient and modern. Let us now take a glimpse into Plato's world—at least into a relatively small but important part of that world, of particular relevance to the nature of physical reality.

Notes

Section 1.2

1.1. Unfortunately, almost nothing reliable is known about Pythagoras, his life, his followers, or of their work, apart from their very existence and the recognition by Pythagoras of the role of simple ratios in musical harmony. See Burkert (1972). Yet much of great importance is commonly attributed to the Pythagoreans. Accordingly, I shall use the term 'Pythagorean' simply as a label, with no implication intended as to historical accuracy.

1.2. This is the pure 'diatonic scale' in which the frequencies (in inverse proportion to the lengths of the vibrating elements) are in the ratios 24 : 27 : 30 : 32 : 36 : 40 : 45 : 48, giving many instances of simple ratios, which underlie harmonies that are pleasing to the ear. The 'white notes' of a modern piano are tuned (according to a compromise between Pythagorean purity of harmony and the facility of key changes) as approximations to these Pythagorean ratios, according to the *equal temperament* scale, with relative frequencies $1:\alpha^2:\alpha^4:\alpha^5:\alpha^7:\alpha^9:\alpha^{11}:\alpha^{12}$, where $\alpha = \sqrt[12]{2} = 1.05946\ldots$. (Note: α^5 means the fifth power of α, i.e. $\alpha \times \alpha \times \alpha \times \alpha \times \alpha$. The quantity $\sqrt[12]{2}$ is the twelfth root of 2, which is the number whose twelfth power is 2, i.e. $2^{1/12}$, so that $\alpha^{12} = 2$. See Note 1.3 and §5.2.)

Section 1.3

1.3. Recall from Note 1.2 that the *n*th power of a number is that number multiplied by itself *n* times. Thus, the third power of 5 is 125, written $5^3 = 125$; the fourth power of 3 is 81, written $3^4 = 81$; etc.

1.4. In fact, while Wiles was trying to fix a 'gap' in his proof of Fermat's Last Theorem which had become apparent after his initial presentation at Cambridge in June 1993, a rumour spread through the mathematical community that the mathematician Noam Elkies had found a counter-example to Fermat's assertion. Earlier, in 1988, Elkies had found a counter-example to Euler's conjecture—that there are no integer solutions to the equation $x^4 + y^4 + z^4 = w^4$—thereby proving it false. It was not implausible, therefore, that he had proved that Fermat's assertion also was false. However, the e-mail that started the rumour was dated 1 April and was revealed to be a spoof perpetrated by Henri Darmon; see Singh (1997), p. 293.

1.5. Technically it is a Π_1-sentence; see §16.6.

1.6. I realize that, in a sense, I am falling into my own trap by making such an assertion. The issue is not really whether the mathematicians taking such an extreme subjective view happen to constitute a tiny minority or not (and I have certainly not conducted a trustworthy survey among mathematicians on this point); the issue is whether such an extreme position is actually to be taken seriously. I leave it to the reader to judge.

1.7. Some readers may be aware of the results of Gödel and Cohen that the axiom of choice is independent of the more basic standard axioms of set theory (the Zermelo–Frankel axiom system). It should be made clear that the Gödel–Cohen argument does not in itself establish that the axiom of choice will never be settled one way or the other. This kind of point is stressed, for example, in the final section of Paul Cohen's book (Cohen 1966, Chap. 14, §13), except that, there, Cohen is more explicitly concerned with the *continuum hypothesis* than the axiom of choice; see §16.5.

Section 1.4

1.8. There is perhaps an irony here that a fully fledged anti-Platonist, who believes that mathematics is 'all in the mind' must also believe—so it seems—that there are no true mathematical statements that are in principle beyond reason. For example, if Fermat's Last Theorem had been inaccessible (in principle) to reason, then this anti-Platonist view would allow no validity either to its truth or to its falsity, such validity coming only through the mental act of perceiving some proof or disproof.

1.9. See e.g. Penrose (1997b).

1.10. My own views on the kind of change in our physical world-view that will be needed in order that conscious mentality may be accommodated are expressed in Penrose (1989, 1994, 1997a,1997b).

2
An ancient theorem and a modern question

2.1 The Pythagorean theorem

Let us consider the issue of geometry. What, indeed, are the different 'kinds of geometry' that were alluded to in the last chapter? To lead up to this issue, we shall return to our encounter with Pythagoras and consider that famous theorem that bears his name:[1] for any right-angled triangle, the square of the length of the hypotenuse (the side opposite the right angle) is equal to the sum of the squares of the lengths of the other two sides (Fig. 2.1). What reasons do we have for believing that this assertion is true? How, indeed, do we 'prove' the Pythagorean theorem? Many arguments are known. I wish to consider two such, chosen for their particular transparency, each of which has a different emphasis.

For the first, consider the pattern illustrated in Fig. 2.2. It is composed entirely of squares of two different sizes. It may be regarded as 'obvious' that this pattern can be continued indefinitely and that the entire plane is thereby covered in this regular repeating way, without gaps or overlaps, by squares of these two sizes. The repeating nature of this pattern is made manifest by the fact that if we mark the centres of the larger squares, they form the vertices of another system of squares, of a somewhat greater size than either, but tilted at an angle to the original ones (Fig. 2.3) and which alone will cover the entire plane. Each of these tilted squares is marked in exactly the same way, so that the markings on these squares fit together to

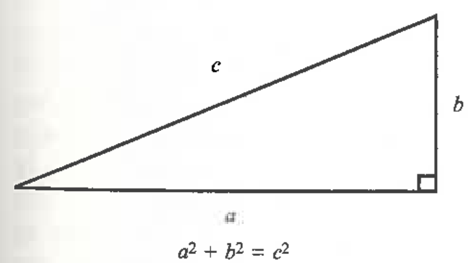

Fig. 2.1 The Pythagorean theorem: for any right-angled triangle, the squared length of the hypotenuse c is the sum of the squared lengths of the other two sides a and b.