1
Introduction

Basic astronomical phenomena

A great deal of human effort has been expended over the past 4000 years or so in trying to predict and explain the motions of the Sun, Moon, planets and stars. Since Babylonian times, this quest has relied heavily on mathematics, and the developments in man's understanding of the heavens have been inextricably linked to progress in the mathematical sciences. As far as the Sun, Moon and planets are concerned, attempts at an explanation of their motion using mathematical techniques began in ancient Greece and the first mathematical model of the heavens was constructed by Eudoxus in the fourth century BC. The final piece of the celestial jigsaw was supplied by Einstein's theory of general relativity in the twentieth century and, since 1915, all major phenomena associated with planetary motion have possessed theoretical explanations. This should not be taken to imply that we know everything about the future positions of the bodies in our Solar System. Indeed, the researches of Poincaré in the late nineteenth century have led us to a much clearer understanding of the limitations of theoretical predictions. Before embarking on the fascinating story that describes the endeavours of men such as Ptolemy, Copernicus, Kepler, Newton, and Laplace, we will begin by familiarizing ourselves with the various heavenly phenomena that people have for so long sought to explain.

For a variety of reasons, early astronomers thought that the Earth was stationary and that the heavenly bodies moved around it. On the face of it this is an extremely natural assumption to make, the evidence to the contrary is far from obvious. The fact that the natural interpretation of the situation is wrong is one of the reasons why astronomy has such an absorbing history. Progress in man's understanding of the nature of the Universe has not been a gradual refinement of simple and intuitive ideas, but a struggle to replace the seemingly obvious by, what to many, was patently absurd. Nowadays, we accept that the idea of a stationary Earth at the centre of the Universe is wrong and that the Earth, Sun, planets, and stars are all in motion relative to each other, but in order to understand early approaches to astronomy it is often helpful to throw away our modern notions and to try and picture the Universe as the ancients would have done.

The most obvious of the objects visible in the sky is the Sun, and systematic observations of its motion across the sky were made by the Babylonian and Egyptian civilizations using a *gnomon*, which was simply a primitive sundial consisting of a stick placed vertically on a horizontal surface.[1] During the course of a day both the length and direction of the *gnomon*'s shadow vary; when the Sun is high in the sky the shadow is short and although the minimum length of the shadow varies from day to day, the direction of the Sun at the time when the shadow is shortest is always the same. For an observer in the northern hemisphere (and we will assume throughout that the observer is in the northern hemisphere) this direction defines due north. Each day, the Sun rises in the eastern part of the sky, travels across the sky reaching its highest point in the south in the middle of the day, and then sets in the western part of the sky. However, the amount of time for which the Sun is visible and its daily path across the sky are not constant, e.g. the points on the horizon at which the Sun rises and sets undergo variations. Observations over a sufficient period of time would show that these variations repeat, and that the period with which they do so is related to the weather (in most parts of the world); in this way one is led to the concept of a year with its seasonal changes in climate.

In Egypt, where the seasons are not particularly noticeable, this period was recognized to be the same as that associated with the flooding of the Nile and, hence, crucial to people's lives. The Egyptians noticed that for part of the year Sirius, the brightest star in the sky, was invisible as it was too close to the Sun, but that the floods began soon after the time when the star rose in the eastern sky just before dawn, and they measured the period between these so-called heliacal risings at a little over 365 days. We call this period the **tropical year**, and it forms the basis of our calendar. For calendrical purposes, the Egyptians used a year of precisely 365 days and the simplicity of such a system was beneficial in the extreme to astronomers; indeed, it was used by Copernicus as late as the sixteenth century. The division of a day into 24 h is also of ancient origin, though originally these were of unequal length and depended on one's location and the time of year in such a way that there were always 12 h of daylight and 12 h of darkness.[2]

[1] The early use of shadow sticks is described in Fermor (1997).
[2] For the purposes of everyday life, the change to 24 equal hours did not take place until the fourteenth century with the advent of the mechanical clock, but hours of equal length were used

The next most obvious heavenly body is the Moon, and this is seen to undergo not only variations in its track across the sky but also in its form, changing within a period of about $29\frac{1}{2}$ days from thin crescent to circular disc (**full moon**) back to thin crescent again and then disappearing for two or three nights (**new moon**). These changes in form are known as the 'phases of the Moon' and are associated with another characteristic length of time, known as a **lunation** or **synodic month**, that formed the bases of many ancient calendars. When the Moon is full, it is opposed diametrically to the Sun, and we say that the Sun and the Moon are in **opposition**. When the Sun and the Moon are in the same direction – which happens at the new moon – they are said to be in **conjunction**. This same terminology is used for any pair of heavenly bodies. Oppositions and conjunctions are known collectively as **syzygies**.

If observers look toward the sky on a clear night, they see a large number of stars that appear to lie on a spherical surface with themselves at the centre; this imaginary surface is called the **celestial sphere**. The stars appear to move over the surface of this sphere but the distance between them remains constant, and it was for this reason that many ancient astronomers regarded the celestial sphere as a real entity with the stars attached physically to it. This physically real celestial sphere had its centre at the centre of the Earth, but it is more convenient to take the imaginary celestial sphere (illustrated in Figure 1.1) as having its centre at the observer O. The point on the celestial sphere directly above the observer is called the **zenith**, Z.

Although we now know that the stars are not all the same distance from the Earth, it is not at all hard to imagine when looking at them that they are equidistant from us, though we have no immediately simple way of determining just how far away they are. Ignoring the question of distance, the celestial sphere gives us an easy way of describing the direction of a star: it is the same as the direction of the line that points from an observer to the point on the celestial sphere where the star appears to lie. We can then talk about the **apparent distance** between two stars as the angle between the lines pointing towards each of them.

Some careful observation will show that the stars move as if they were attached to the celestial sphere and it was rotating about an axis that intersects the sphere at a point in the northern sky. This point, which nowadays is close to the star Polaris, is called the **north pole**, P_{N}, of the celestial sphere and we can easily imagine that there is a **south pole**, P_{S}, in the part of the sky invisible to us.

from ancient times in the context of scientific discussions. A division of the day into equal temporal units was used in China from the second century BC. See Dohrn-van Rossum (1996) for a complete history of the hour.

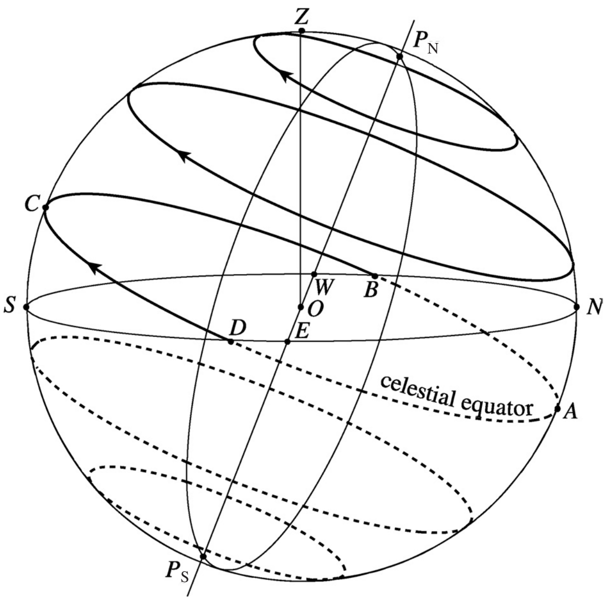

Fig. 1.1. The celestial sphere.

The stars move around the north pole in an anticlockwise direction with some stars always above the **horizon** ($NWSE$), so-called **circumpolar** stars (or, as some ancient observers termed them, 'those that know no weariness'),[3] and some dipping below the horizon as indicated by the dashed lines in Figure 1.1. We can imagine easily that there are other stars that remain invisible below the horizon. The circle $ABCD$, which lies in a plane through O midway between the poles, is called the **celestial equator**. If we were at the north pole of the Earth, then the north pole of the celestial sphere would correspond to the zenith and all visible stars would be circumpolar, whereas if we were on the equator, the north pole of the celestial sphere would be on the horizon and there would not be any circumpolar stars. For a general point in the northern hemisphere, the angle between OZ and OP_N is 90° minus the latitude of the observer. The celestial sphere completes 1 revolution in about 1 day, and this rotation is called the 'daily' or **diurnal motion**. The Moon also participates in this daily rotation

[3] It is these circumpolar stars that are most suggestive of the spherical nature of the heavens. Ptolemy, in AD second century stated explicitly that these stars were instrumental in leading astronomers to the concept of a celestial sphere (PTOLEMY *Almagest*, Book I, 3 (see Toomer (1984)).

as does the Sun, though since the stars and the Sun are not seen at the same time, this is harder to appreciate.

Provided we have an accurate means of measuring time, we can observe that the stars actually complete a revolution about the pole in about 23 h 56 min, so that they return to the same place at the same time in 1 year.[4] For an ancient observer, this observation was more difficult. However, by observing stars rising above or setting below the horizon at about sunrise or sunset, one notices that the stars are moving faster than, and therefore gradually changing their position with respect to the Sun, returning to the same position after 1 year. If we regard the stars as fixed on the celestial sphere, then the Sun must move relative to the stars in the direction opposite to the diurnal motion (i.e. from west to east), completing one circuit of the celestial sphere in a year. There are other features of the motion of the Sun that need to be explained. For example, the points on the horizon at which the Sun rises and sets vary from day to day, as does the midday height of the Sun and, indeed, the length of time the Sun is above the horizon. A great deal of careful observation led ancient astronomers to the conclusion that the complex motion of the Sun was built up from two much simpler motions: the first was the daily rotation of the celestial sphere and the second was a much slower **annual motion** that took place on an oblique great circle ($AQCR$ in Figure 1.2). This circle is called the **ecliptic**, the reason being that for eclipses to occur, the Moon must be on or near it, and the angle at which it cuts the celestial equator is called the **obliquity** of the ecliptic, ε. Nowadays, the value of ε is about $23^\circ\,27'$, but the obliquity actually decreases very slowly with time. In 3000 BC, it was about $24^\circ\,2'$.[5]

In order to describe the position of a body on the celestial sphere, we need some sort of coordinates. One possibility, introduced by Babylonian astronomers, is to measure angles relative to the ecliptic. The **ecliptic longitude** measures the distance around the ecliptic, whereas the **ecliptic latitude** measures the distance north or south of this line. Hence, by definition, the ecliptic latitude of the Sun is 0°. A common alternative is to use a system based on the celestial equator, and here the distance around the equator is known as the **right ascension**, whereas the distance north or south of the equator is the **declination**.

From Figure 1.1, it can be seen that when the Sun lies on the celestial equator the lengths of the day and night are equal (since the arcs DCB and BAD are the same length) and, hence, the points A and C in Figure 1.2 are known as the **equinoctial points**, and the times when the Sun is at these points are the **equinoxes**. The time at which the Sun is at A (i.e. when it crosses the celestial

4 Twenty-four hours divided by 365.25 (the number of days in 1 year) is just under 4 min.
5 Thurston (1994).

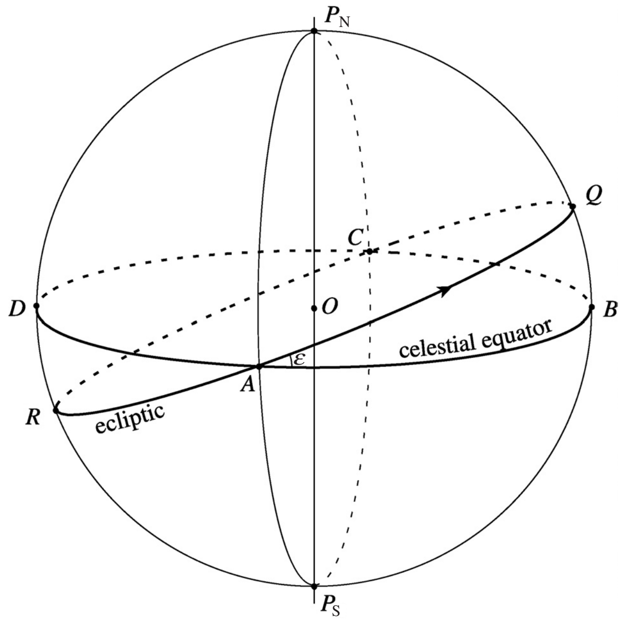

Fig. 1.2. The annual motion of the Sun.

equator from south to north) is the **spring** or **vernal equinox**, whereas the Sun is at C at the **autumnal equinox**. Ancient astronomers were aware of the fact that the Sun does not move around the ecliptic at a uniform rate – the autumnal equinox is about 186 days after the vernal equinox, but only 179 days pass before the autumnal equinox is reached again.

The Sun is at point Q – its most northerly extreme – at the **summer solstice**, and at its most southerly extreme, R, at the **winter solstice**. The points Q and R are known as the **solstitial points**. Observations over a short period of time suggest that the equinoctial and solstitial points remain fixed with respect to the fixed stars, but actually they do not; the equinoctial points rotate around the celestial equator with a period of about 26 000 years. Surprisingly, perhaps, this phenomenon, called the ‘precession of the equinoxes’, was recognized as early as the second century BC. The fact that the equinoxes precess means that the time taken for the Sun to return to the same position on the ecliptic (the tropical year) is different from the time taken to return to a fixed star. This latter period is known as the **sidereal year**. The modern values for these periods are approximately 365.242 days for the tropical year and 365.256 days for the sidereal year. When we use the term ‘year’ without qualification we mean the tropical year, though in most situations the difference is unimportant.

Table 1.1. *The periods of the moon.*

Month	Length in days
synodic	29.531
sidereal	27.322
draconitic	27.212
anomalistic	27.555

Observations of the Moon show that its monthly path round the celestial sphere is also a great circle that is very close to the ecliptic but inclined slightly to it. The orbit of the Moon crosses that of the Sun at two points called 'the **nodes** of the orbit' and the straight line joining these points is called the **nodal line**. At one of its nodes, the Moon is moving from south of the ecliptic to north of it, and we call this the **ascending node**, the other point of crossing being the **descending node**. The period of the rotation of the Moon around the celestial sphere is known as the **sidereal month**, which is, of course, different from the synodic month because that measures the motion of the Moon with respect to the Sun, which itself is moving around the ecliptic. There are two other important periods associated with the Moon which were recognized in ancient times. First, there is the **draconitic month**, which is the period between successive ascending (or descending) nodes,[6] or equivalently the time it takes for the Moon to return to the same distance from the ecliptic (i.e. to the same ecliptic latitude) while travelling in the same direction. Second, the speed with which the Moon moves across the sky relative to the stars is variable (varying between about 12 and 15° per day) and the period of time it takes for the Moon to return to the same speed is called the **anomalistic month**. The modern mean values for the various periods of the Moon are shown in Table 1.1, though the actual values for any given month may differ from these by as much as 7 h.

Perhaps the most dramatic of celestial events which are easily observable are **eclipses**, and these are of two types: lunar and solar (see Figure 1.3).[7] In a **lunar eclipse** the Earth passes between the Sun and the Moon and casts its shadow over the Moon's surface, whereas in a **solar eclipse** it is the shadow

[6] In medieval times, the part of the Moon's orbit south of the ecliptic was known as the 'dragon' (which devoured the Moon during eclipses) and from this we get the terminology 'dragon's head' for the ascending node and 'dragon's tail' for the descending node. An example of this usage can be found in Chaucer's *Treatise on the Astrolabe* (1391), one of the oldest surviving scientific works written in English. The periods between successive nodes has, over time, been termed the dracontic, draconic and draconitic month, the words deriving from the Greek for 'dragon'.

[7] Eclipses can be categorized further depending on the precise arrangement of the three bodies (see, for example, Payne-Gaposchkin (1961)).

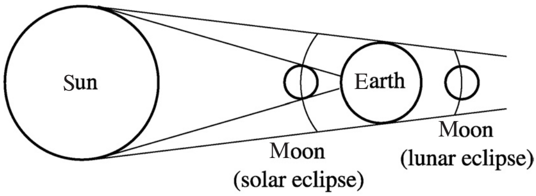

Fig. 1.3. The nature of eclipses.

of Moon that passes over the Earth. Thus, a lunar eclipse takes place when the Moon is in opposition to the Sun (i.e. when it is full) and simultaneously at (or near) one of its nodes. Similarly, a solar eclipse occurs when the passage of the Moon corresponds, through one of its nodes, to a conjunction with the Sun. Now, if the nodal line of the orbit of the Moon had a fixed orientation with respect to the stars, one of its nodes would lie directly between the Earth and the Sun every half year. Actually, the nodal line rotates (a fact known to the ancient Greeks) and as a result a node lies on the Earth–Sun line once every 173.3 days. A solar eclipse will occur if there is a new moon close enough to this time. It turns out that the shadow of the Moon will touch the Earth if the new moon appears within about $18\frac{3}{4}$ days either side of the alignment of a node and, thus, there is a $37\frac{1}{2}$-day eclipse season every 173 days during which a solar eclipse may be visible.

A similar analysis applies to lunar eclipses, but since the shadow cone of the Earth narrows as one moves further from the Sun, the full moon needs to appear nearer to the point of alignment than the $37\frac{1}{2}$-day window that exists for solar eclipses. As a result, lunar eclipses actually are less frequent than the solar variety. In any one calendar year, there are between two and five solar eclipses, but there can be no more than three lunar eclipses per year and there might not be any. The reason solar eclipses are seen so much more rarely than lunar ones is, of course, due to the fact that the area of the Earth covered by the shadow of the Moon in a solar eclipse is very small, and so a particular solar eclipse is visible from only a small part of the surface of the Earth. On the other hand, the shadow of the Earth can block out the light of the Sun from the whole of the surface of the Moon, and such a lunar eclipse will be visible from anywhere on the Earth from which the Moon normally would be visible.

Eclipses of both the Sun and Moon were observed by ancient astronomers and would no doubt have aroused great interest. Over a period of time it would have become clear that eclipses of the Moon occur only when the Moon is full, and that eclipses of the Sun occur only at new moon. The fact that solar eclipses are caused by the Moon passing in front of the Sun would not have been hard to appreciate, but the fact that eclipses of the Moon are caused by the shadow of

Table 1.2. *The zodiacal and synodic periods of the planets.*

	Zodiacal period (years)	Synodic period (days)
Mercury	1	115
Venus	1	584
Mars	1.88	780
Jupiter	11.86	399
Saturn	29.46	378

the Earth passing over it requires a rather deeper understanding of the situation and was probably not realized until much later. The reasons behind both types of eclipse were, however, fully understood by the time of the ancient Greeks.

Ancient astronomers were also aware that five of the star-like objects in the sky changed their position relative to the other stars. These five objects – now named after the Roman gods Mercury, Venus, Mars, Jupiter and Saturn – are the **planets**, from the Greek for 'wanderer'. Careful observations of these objects reveal that, like the Sun, as well as participating in the daily rotation of the heavens, they, too, move around the celestial sphere though with differing periods, and also that while they move predominantly in the same direction as the Sun – from west to east – they sometimes switch back and, for a time, move from east to west in so-called **retrograde** motions. The periods of retrograde motion are linked to the motion of the Sun – the centre of the retrograde motion for Mars, Jupiter and Saturn always occurring when the planet is in opposition to the Sun – whereas for Mercury and Venus this phenomenon occurs at conjunction. The planets also remain close to the ecliptic, the maximum deviation for any of them being 8°, and thus all the wandering heavenly bodies can be found within a strip on the celestial sphere 16° across centred on the ecliptic. This strip, therefore, is very important, and is known as the **zodiac** and was divided by the Babylonians into twelve equal parts: the signs of the zodiac. The average time it takes for a planet to complete 1 revolution around the ecliptic is its **zodiacal period**, and the average period between successive occurrences of retrograde motion is known as the **synodic period** of the planet; for the five planets visible with the naked eye these are given in Table 1.2.

The fact that the Sun, Moon and planets are nearer the Earth than the stars may have been suggested by **occultations**, the temporary disappearance of one heavenly body behind another. The most obvious example is the Moon, which sometimes eclipses the Sun and also is easily observed to pass in front of the planets and stars. Thus, all ancient astronomers agreed that the Moon was the closest of the heavenly bodies. This view was, of course, supported by the Moon's enormous size compared to the other planets and by the fact that various details could be made out on its surface. However, there is no easy way of determining how far away each celestial body is, and so ancient astronomers took the view that their distance probably was related to the speed with which they traversed the celestial sphere. This was consistent with the fact that the Moon is the swiftest of the heavenly bodies.

The Sun's journey around the ecliptic takes, by definition, 1 year, and it was found that Mars returned to the same place among the stars on average after about 2 years, Jupiter after 12 years and Saturn after $29\frac{1}{2}$ years. These planets, being slower than the Sun, were therefore thought to be further away and, hence, higher in the sky, and were termed the **superior** planets. On the other hand, Venus and Mercury both complete 1 revolution about the Earth in the same average time (1 year) and there was considerable disagreement about the correct ordering of these two planets. Eventually, a consensus was reached that they were closer to the Earth than the Sun, and they became known as the **inferior** planets. The observed behaviour of the superior and the inferior planets is different in another important way: Mercury and Venus are never seen far from the Sun, the maximum differences in longitude being about 29 and 47°, respectively, and so they are only ever visible near dawn or sunset, whereas the difference in longitude between the Sun and Mars, Jupiter or Saturn can take any value and they can be visible at any time of night.

The Sun, Moon and five planets, together with the fixed stars, were the only heavenly bodies recognized in antiquity, and this situation did not change until Galileo pointed a telescope at the night sky in 1609–10 and discovered the moons of Jupiter. No further planets were discovered until the eighteenth century, and comets were not considered as celestial bodies until the pioneering observations of Tycho Brahe in the late sixteenth century (prior to this they were thought of as atmospheric phenomena).

The problem for astronomers, then, was to explain the phenomena that have been described above. In the beginning, attempts were limited to qualitative explanation, but the models that were developed could not produce accurate quantitative predictions for astronomical events. The quantitative problem is much, much harder, and it exercised many of the greatest minds over a period of more than 2000 years, from Babylonian times to the twentieth century.

Babylonian astronomy

The heavenly phenomena were of great importance to the Babylonians, as they were perceived as omens and just about every possible astronomical event had some significance. For example, when it came to the retrograde motion of the planets it was not simply the retrograde motion itself, but also where it took place with respect to the stars, that was important:

> When Mars comes out of the constellation Scorpius, turns and reenters Scorpius, its interpretation is this: . . . do not neglect your guard: the king should not go outdoors on an evil day.[8]

In order to improve their abilities to predict such phenomena, the Babylonians developed a tradition of observing and recording celestial events; examples of Babylonian astronomical observations exist which date back as far as 1600 BC. These are not particularly accurate, but show that the practice of observing and then recording the results has a long history in Babylonian culture. Scribes systematically began documenting celestial phenomena (e.g. eclipses) in about the eighth century BC, and over the next 700 or 800 years produced a large quantity of data that would later prove invaluable to Greek astronomers.[9] In order to carry out their work, astrologers needed tables of the future positions of heavenly bodies (which we now call 'ephemerides') and this desire was the driving force behind the production of such tables for over 2000 years.

The Babylonian number system was sexagesimal (base 60) and positional.[10] The reason for the use of 60 is unclear, but it may well have been because 60 is divisible exactly by lots of small integers, so many calculations can be done without the use of fractions. Nevertheless, because of the positional nature of the system, the treatment of fractions was vastly superior to the methods employed in Egypt and Greece, and as a result it was used by virtually all Western astronomers up until the sixteenth century, when decimal fractions began to take over. The superiority of the positional system of numeration helps to explain why quantitative astronomy reached a much greater level of sophistication in Babylonia than in other contemporary civilizations. The influence of the sexagesimal number system is still apparent in our measurement of time and angles, with 1 h being divided into 60 min, and so on. The Babylonians are

[8] This is just one of many such examples described in Swerdlow (1998a).

[9] The earliest Babylonian observations that the later Greek astronomers were able to use were made during the reign of Nabonassar (747–33 BC) and thus this period became a key reference point in the Greek system of reckoning time. Babylonian astronomers are often referred to as Chaldeans, a tribe from the southern part of the area known as Babylonia that established a new dynasty in 625 BC. Our knowledge of Babylonian astronomy was transformed during the twentieth century with the deciphering of hundreds of astronomical documents. The way this new information was absorbed into mainstream history of science is described in Rochberg (2002).

[10] We will follow standard modern practice of writing sexagesimal numbers with a semicolon representing the sexagesimal point. Thus, $1,\ 24;\ 20 = 60 + 24 + \frac{20}{60} = 84\frac{1}{3}$.

also responsible for dividing the circle up into 360 equal parts, which we call 'degrees', though this would not appear to have been a direct consequence of their sexagesimal system, but seems to have been due to the length of the year being about 365 days, and so in 1 day the Sun moves about 1° with respect to the stars.

The Babylonians began to turn their observational records into a mathematical theory around 500 BC. It was also around this time that they created the concept of the zodiac, dividing the strip straddling the ecliptic into twelve equal subdivisions of 30° each, so as to aid computations. From their data they realized that there were fixed periods of time associated with many astronomical events and, since knowledge of these events had great benefits in terms of prognostication, they spent considerable time and effort determining these periods accurately. Typically, they would develop relationships in which m intervals of one type were equated with n different intervals. For example, they discovered the approximate equivalence between 19 years and 235 synodic months; a relationship that formed the basis of their luni-solar calendar and is now known as the 'Metonic cycle' after the Athenian who tried unsuccessfully to incorporate it into the Greek calendar (since $235 = 19 \times 12 + 7$, years consisted of 12 months, with 7 intercalary months every 19 years). Another example is the Venus cycle in which the planet completes five synodic periods in 8 years ($8 \times 365 = 5 \times 584$).

The Babylonians developed sophisticated mathematical techniques that enabled them to develop an accurate predictive astronomy.[11] Most Babylonian astronomy was concerned with the motion of the Moon (e.g. determining when the new moon would first become visible, since this was the basis of their calendar), but they also tabulated planetary oppositions and conjunctions. Our knowledge of their astronomy comes largely from ephemerides, which list the positions of the Moon or planets over a number of years at regularly spaced intervals. The values listed are the results of calculations rather than of observations, so they reveal the underlying mathematical techniques that the Babylonians used to model the heavenly phenomena.

Fundamental to the phenomena the Babylonians were interested in was the motion of the Sun, and from their ephemerides it has been possible to deduce how they modelled the Sun's non-uniform motion. Interestingly, there were two methods, both of which were in use during the whole period for which ephemerides have been found (the last three centuries BC). The simpler version, known as 'System A', implied a motion for the Sun of 30° each mean

[11] The summary given here is based on the excellent introduction to Babylonian astronomy given in Neugebauer (1969).

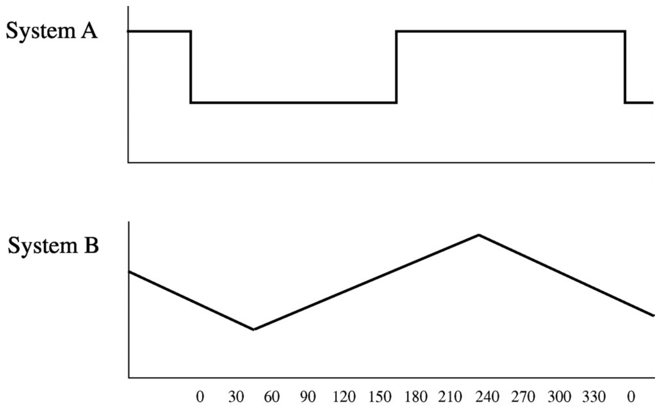

Fig. 1.4. Variations in the solar motion in Babylonian astronomy.

synodic month for just over half the year and then $28; 7, 30^\circ$ for the remainder. Calculations reveal that this system corresponds to the relation

$$1 \text{ year } = 12;\ 22,\ 8 \text{ synodic months.}$$

In 'System B', rather than have the Sun's motion change abruptly at two points during the year, it was made to vary continuously, oscillating between the maximum and minimum values of $30;\ 1,\ 59^\circ$ and $28;\ 10,\ 39,\ 40^\circ$ per mean synodic month. In order to represent this, the Babylonians used alternate increasing and decreasing arithmetic progressions that are best illustrated as a zigzag function. The relation between the length of the year and that of the synodic month was the same as in System A. Both systems are illustrated in Figure 1.4, in which the numbers denote the ecliptic longitude of the Sun measured from the vernal equinox. Ephemerides also exist that concern the daily motion of the Moon in which its speed is treated in a similar way by means of a zigzag function.

The planetary phenomena are more complex, but the mathematical techniques with which the Babylonians attempted to model them were essentially the same as the ones they used to represent the Sun's motion.[12] For the case of a superior planet, there were five characteristic phenomena that concerned Babylonian astronomers, and these are shown schematically in Figure 1.5. When the planet is in conjunction, it cannot be seen because of the Sun's glare but as it moves around the ecliptic relative to the Sun there comes a point, Γ, where it first becomes visible just before dawn (its heliacal rising). Similarly, the planet

[12] Detailed descriptions of the technicalities of Babylonian astronomy can be found in, for example, Aaboe (1958, 1964, 1980), Swerdlow (1998a, 1999), Steele (2000, 2003).

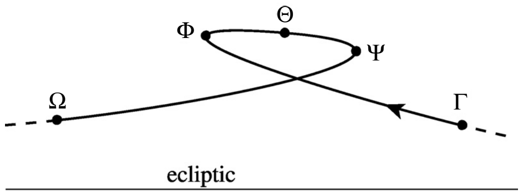

Fig. 1.5. The characteristic phenomena for a superior planet in Babylonian astronomy.

eventually will disappear as it approaches conjunction again, and its last appearance just after sunset (its heliacal setting) is labelled Ω. In-between, the planet will undergo a period of retrograde motion around its opposition, Θ, and the two positions at which its motion along the ecliptic is stationary, Φ and Ψ, make up the points of interest. The main problem of Babylonian planetary astronomy was, given the longitude of one of these phenomena for a particular planet, to predict the longitude of the next occurrence of the same phenomenon. It was these differences in longitude that the Babylonians attempted to model with their zigzag (System B) or step (System A) functions.

As well as developing fairly sophisticated techniques with which to model the periodic phenomena they observed in the heavens, the Babylonians achieved a high level of skill in numerical techniques. What they do not appear to have done, however, was seek to explain what went on in the heavens in terms of a geometrical or physical model – their approach was entirely arithmetical. The desire to construct a model of the Universe, a cosmology, which corresponds to the observed celestial phenomena has its origins in the works of ancient Greek philosophers. However without the accurate quantitative astronomy that the Babylonians created, and the vast amount of data they catalogued, Greek astronomers would not have been able to achieve what they did.

Early Greek astronomy

Early cosmologies were based heavily on man's earthly experiences and owed little to accurate astronomical observation. They fulfilled a well-documented psychological need, providing a stage for the drama of daily life, for the actions of the gods and also supplying meaning to man's existence. Before the ancient Greek civilization, cosmologies were constructed that made no attempt to explain any but the most rudimentary of astronomical phenomena. For example, the Egyptians explained the motion of the Sun by having the Sun god Ra take a daily trip through the air and then each night make a passage through the water. The Moon, on the other hand, was attacked on the fifteenth day of each month by a sow and after 2 weeks of agony the Moon died and was reborn.

Gradually, as the Greek civilization progressed, cosmologies became more sophisticated, and the idea that the way to find out about the nature of the Universe is through accurate astronomical observation, now a well-established tenet of Western thought, developed. One of the earliest examples we have showing Greek interest in the heavens comes from the writings of Homer in which the Earth is a flat circular disc surrounded by the Ocean river and covered by the vault of heaven. The Homeric epics show that in the eighth century BC the Greek understanding of the natural world was very primitive, and their awareness of astronomical phenomena was considerably less than that of their Babylonian contemporaries. Further evidence of early Greek astronomical knowledge can be obtained from Hesiod's poem 'Works and Days', written about a century after Homer's time, which among other things, contains what might be described as an agricultural calendar. Hesiod described how the seasons are related to the heliacal risings and settings of certain stars and, in contrast to Homer, connected astronomical events to the lives of ordinary people.

Up until about 300 BC, Greek astronomy was almost entirely qualitative. Indeed, there is no evidence that the accurate prediction of heavenly phenomena was even thought of as a desirable goal. This situation changed when the Greeks came into contact with the quantitative methods of Babylonian astronomers during the expansion of their empire under Alexander the Great. Before then, however, celestial phenomena were the subject of a great deal of philosophical debate that provided the basis on which later astronomers could build. Four major schools of philosophical thought existed during the 300 or so years prior to the construction of the first mathematical model of the Universe by Eudoxus.[13] There were the Ionians, a group founded by Thales of Miletus (in Asia Minor, modern-day Turkey) in about 600 BC. During the sixth century BC the region of Asia Minor went through a considerable upheaval due to the expansion of the Persian empire and many philosophers travelled to other parts of the Greek empire. One member of the Ionian School, Xenophanes of Colophon, migrated to southern Italy and eventually settled in Elea, which became an important philosophical centre. Perhaps the most famous of the early Greek philosophical schools was that set up by Pythagoras (who is said to have been taught by Thales), again in southern Italy, and finally there was the celebrated school centred around Plato's Academy in Athens. It should always be borne in mind

[13] For a modern introduction to ancient Greek philosophy, see Kenny (1998).

that it is difficult to be certain about the specific theories espoused by these early thinkers, since most of our knowledge comes from remarks of later (and not always reliable) authors.

The Ionian philosophers began the process of determining the nature of things but did not progress very far towards a rational description of the Universe. We know very little about the people who made up this School and none of their writings have survived. What little we do know is due to Greek historians and commentators. As far as astronomy was concerned, the philosophers of the Ionian School thought the Earth was flat, and they had a very poor understanding of the nature of the Sun and Moon.

According to Aristotle, the founder of Ionian natural philosophy was Thales, whose ideas (cultivated during travels around Egypt, the Mediterranean and Near East) were built around the fundamental belief that water is the essence of all things. The significance of this is the suggestion that there is an underlying unity to physical phenomena and, hence, that nature is not quite as haphazard as our senses would have us believe. In terms of the structure of the Universe, however, Thales' ideas were about as primitive as those found in Homer. Later authors have ascribed knowledge of the sphericity of the Earth and of the causes of eclipses to him, though it is extremely doubtful that Thales actually possessed this knowledge. He is also credited with predicting a solar eclipse in 585 BC though this is probably false also.[14] The sayings which are attributed to him show us that over 2500 years ago Greek intellectuals were attempting to understand concepts such as space and time:

> Of all things that are, the most ancient is God, for he is uncreated; the most beautiful is the universe, for it is God's workmanship; the greatest is Space, for it contains everything; the swiftest is Mind, for it speeds everywhere; the strongest is necessity, for it masters all; and the wisest Time, for it brings everything to light.[15]

Thales also holds a significant place in the history of mathematics. He is reputed to have been the first person to *prove* geometrical theorems and a number of such proofs are attributed to him, though their nature is unclear. Certainly, it is true that the use of logical proofs in mathematics was developed over the three centuries before Euclid wrote the *Elements* in about 300 BC, and that later Greek writers credited Thales with inventing this deductive method, but the precise nature of Thales' contribution is uncertain.

Perhaps the most important of the Ionian philosophers was Anaximander, a pupil of Thales. Almost all we know of his doctrines comes ultimately from Theophrastus' *Physical Opinions*. In Anaximander's system, the essence of all

[14] See Neugebauer (1969), though an explanation of how Thales might have achieved this feat is given by Hartner (1969). See also Aaboe (1972).
[15] Quoted from Heath (1932).

things was not water or any other physical quality, but the infinite, and there are similarities between Anaximander's 'infinite medium' and the concept of the ether, that played a significant role in science right up to the twentieth century.[16]

He described the Sun, Moon and stars as hollow wheels full of fire, each of which had a small hole in it through which the fire was visible. These narrow openings could close, thus accounting for phenomena such as eclipses or the phases of the Moon. The Earth, about which the wheels rotate, was a flat disc that he thought of as the top surface of a cylinder, suspended freely in space, the height of which was for some reason taken to be one-third of the diameter of the disc.[17] Anaximander also speculated on the sizes and distances of the Sun and Moon, though with no apparent basis for his conjectures:

> the sun is a circle twenty-eight times the size of the earth; it is like a chariot-wheel, the rim of which is hollow and full of fire, and lets the fire shine out at a certain point in it through an opening like the nozzle of a pair of bellows: such is the sun.[18]

The significance of Anaximander's ideas lies much less in his conclusions than in the questions he was trying to answer: how big are the stars?, how far away are they? Perhaps most significant of all, in his cosmology Anaximander replaced the actions of gods with mechanisms familiar to those on Earth.

Anaximander's ideas were extended by Anaximenes, who was about 20 or 30 years younger than Anaximander, though his cosmology remained very primitive. Perhaps wishing to make philosophy more tangible, Anaximenes made air the essence of things:

> Anaximenes of Miletus, son of Eurystratus, who had been an associate of Anaximander, said, like him, that the underlying substance was one and infinite. He did not, however, say it was indeterminate, like Anaximander, but determinate; for he said it was air.[19]

Anaximenes, like Thales before him, believed that all things were alive and that, just as air sustains human life, a different form of air sustains the life of all things in the Universe. He believed that the flat disc of the Earth was carried by the air, as were the flat discs of the Sun and Moon, while the stars were fixed

[16] Anaximander used the Greek word *apeiron* and his meaning was somewhat ambiguous; see Toulmin and Goodfield (1965), p. 66. Bochner (1966) describes *apeiron* as 'some kind of stuff (perhaps so tenuous as to be little more than "spatiality") of which the universe was composed'.

[17] According to Heath (1932), p. xxiii, Anaximander said that the wheels of the Sun and the Moon 'lie obliquely', indicating that he was in some sense aware of the ecliptic and its obliquity to the celestial equator. However, Eudemus, in his lost *History of Mathematics* written in the fourth century BC, attributes this discovery to Oenopides in the fifth century BC (Heath (1913), Chapter XIV). There is no evidence that Oenopides made any measurements of the obliquity, but he did calculate the length of the year as $365\frac{22}{59}$ days.

[18] Aëtius *De placitis*, Book II, Volume 20, 1. Translation from Heath (1932).

[19] Simplicius *Physics.* Translation from Heath (1932).

to the crystalline celestial vault. The heat of the Sun came from the speed of its motion, the stars providing no heat due to their great distance.

The Ionian philosophers asked many interesting questions, most importantly: what is the nature of things?, what is the Sun?, what is the Earth? While their attempts at answers may seem slightly comical, their curiosity was followed up by the greatest minds of antiquity, with impressive results. Modern science grew out of the pioneering spirit of people such as Thales, Anaximander, and Anaximenes.

The most important of the Eleatic philosophers was Parmenides, who was probably a pupil of Xenophanes, who left Ionia for southern Italy in 545 BC, during the time of the Persian conquest of Asia Minor. Xenophanes' most profound thoughts concerned the philosophy of religion, and he believed that certainty of knowledge was unattainable, but he also had some interesting, but still very primitive, ideas about the nature of heavenly bodies. He wrote that the Sun was born again each time it rose and that eclipses occurred due to the light of the Sun being put out. He also believed that the Moon shone by its own light. In Xenophanes' philosophy, the basic element was not water or air, but earth, and he claimed that the world was evolved from a mixture of earth and water and that the earth gradually would be dissolved again by moisture. This process would then repeat in cycles.

Aristotle's pupil Theophrastus credited Parmenides with the significant discovery that the Earth is spherical, but others credit the Pythagoreans with this observation.

> Further we are told that Pythagoras was the first to call the heaven the universe, and the earth spherical, though according to Theophrastus it was Parmenides, and according to Zeno it was Hesiod.[20]

It is, of course, not unlikely that the idea of a spherical Earth occurred independently to a number of people, observant Greek navigators, for example. Parmenides conceived of a Universe consisting of a number of concentric layers centred on the spherical Earth. Here we see for the first time the idea of a system of geocentric spheres that was to play such an important role in the development of future astronomical theories.

The Ionian and Eleatic thinkers represent the origins of Western philosophy and, although their thoughts on the nature of the Cosmos helped stimulate debate, their direct influence on the history of astronomy was slight. The Pythagoreans, on the other hand, had a much more direct bearing on future astronomical thought. Pythagoras – a mystical figure who was born in the first

[20] Diogenes Laërtius *Lives of Philosophers*, Book VIII, 48. Translation from Heath (1932).

half of the sixth century BC – left Asia Minor and founded a School in southern Italy which over a period of about 200 years produced a wealth of influential ideas in astronomy and mathematics. Almost nothing is known of Pythagoras himself and, since the group he founded was very secretive, much of their output is attributed simply to the Pythagoreans as a whole rather than to any one individual. According to Aristotle:

> . . . those who are called Pythagoreans, taking up mathematical things, were first to promote these, and having been reared on them, they supposed that the sources of them were the sources of all things.[21]

Thus, it was mathematics that held the key to understanding the Universe. To the Pythagoreans, mathematics could be broken down into four distinct branches, which in the Middle Ages became known as the *quadrivium*: arithmetic, geometry, music and astronomy. Arithmetic and geometry represent pure mathematics, and music and astronomy are applied mathematics, musical harmony being an application of arithmetic while astronomy is an application of geometry. Like the earlier Ionian philosophers, the Pythagoreans sought to understand the essence of things, but they found their answer in numbers (positive integers). To them, numbers were the basis of all physical phenomena. The Pythagoreans discovered that the notes obtained by plucking two strings sounded harmonious if the lengths of the strings were in ratios of simple whole numbers. For example, a string twice the length produces a note an octave higher, while if the ratio is 3 : 2 we get a fifth.

We learn from Aristotle how the notions of harmony and number were also applied to the study of the heavens:

> . . . they assumed that the elements of numbers were the elements of all things, and that the whole heaven was a harmony and a number. And as many things among numbers and harmonies as had analogies to the attributes and parts of the heavens and to the whole cosmic array, they collected and fit together.[22]

Ratios of whole numbers could be found among the motions of the heavenly bodies and so the Pythagoreans believed that they emitted harmonious sounds – this is the origin of the phrase ‘harmony of the spheres’. The idea that a mathematical structure underlies the Universe originated with the Pythagoreans, and has been extremely influential ever since. It represents one of the cornerstones of modern science.

In terms of astronomical phenomena, the Pythagorean School is generally credited with knowledge of the sphericity of the Earth, and also with some of

[21] Aristotle *Metaphysics*, Book *A*5, 985b. Translation from Aristotle (1999).
[22] Aristotle *Metaphysics*, Book *A*5, 986a. Translation from Aristotle (1999).

the consequences of this, e.g. the existence of regions of the Earth's surface with days and nights each lasting half a year. As we have mentioned, most discoveries of the Pythagoreans are not attributed to any one person, but in the field of astronomy one particular idea has come down to us with the name of its originator attached. Philolaus of Tarentum developed a doctrine in which the centre of the Universe is a fire called *Hestia* – the goddess of the sacred fireplace of homes and public buildings. The spherical Earth, like the other planets, described a circle about this fire with the uninhabited side always turned toward it, completing 1 revolution each day. Outside the Earth were the spheres of the Moon, Sun, Venus, Mercury, Mars, Jupiter, Saturn and the fixed stars. He thought of the Sun as a transparent globe receiving its light and heat from the central fire and from the outside of the heavens. To balance his system he invented a counter-Earth which is always positioned on the opposite side of the central fire, bringing the total number of spheres of heavenly bodies to the sacred number ten. Philolaus' hypothesis had to be abandoned when travellers expanded the horizons of the known world beyond Gibraltar in the west and to India in the east, but still saw no signs of the central fire or counter-Earth.

The ideas that the Earth was not the centre of the Universe and that it was in continuous motion were very bold ones, contrary to all prevailing views of the time and, indeed, for the next 2000 years. It would appear that these views received little support from anyone outside the Pythagorean brotherhood. On the other hand, the idea that the universe was made up of bodies moving in uniform circular motions around a fixed centre was to dominate astronomical thought until the work of Kepler in the seventeenth century. In AD first century Geminus wrote:

> It is presupposed in all astronomy that the sun, the moon, and the five planets move in circular orbits with uniform speed in a direction opposite to that of the universe. For the Pythagoreans, who were the first to apply themselves to investigations of this kind, supposed the motions of the sun, the moon and the five planets to be circular and uniform . . . For which reason, they put forward the question as to how the phenomena might be explained by means of circular and uniform motions.[23]

After the defeat of the Persians in 479 BC, Athens became an extremely wealthy city, attracting a number of intellectuals. One of the first to move there, and to bring with him the knowledge of Ionian natural philosophy, was Anaxagoras of Clazomenae, who became a friend of the city's ruler Pericles,

[23] Quoted from Goldstein and Bowen (1983). In older books, and also in the *Dictionary of Scientific Biography*, Geminus is said to have lived in the first century BC, but this is now thought to be an error caused by a confusion between ancient calendars (see Neugebauer (1975), pp. 579–81).

and devoted his life to investigations of the heavens. He proclaimed Mind as the moving principle in the Universe. Many people before him had speculated on the origin of the light from the Moon, but Anaxagoras was the first to state clearly that the cause was the reflection of light from the Sun, which he thought of as a vast mass of incandescent metal, and he understood that lunar eclipses occur when the Earth blocks the illumination of the Sun. He was the first to think of the seven wandering heavenly bodies in the order Sun, Moon, followed by the five other planets, an arrangement adopted by a number of later astronomers.

When Pericles' popularity began to dwindle, Anaxagoras, his protégé, became vulnerable and his speculations about the sizes of the Sun and Moon (that the Moon was as large as the Peloponnese – an area of mainland Greece – and the Sun was larger!), and those on the causes of Earthly phenomena and the nature of the heavens (he maintained that they were of the same general nature), led to him being convicted of impiety. He died in exile.

During the fourth century BC a number of centres for advanced learning were set up in and around Athens, by far the most influential being the Academy set up by Plato, who was a pupil of the philosopher Socrates. While not a mathematician himself, Plato was influenced strongly by Pythagorean doctrines. He believed that abstract mathematical concepts had a real, independent, existence, and that true knowledge could be obtained only through study of these so-called Platonic forms. The natural world was to be understood via unchanging mathematical laws, untarnished by the imperfections of our senses. His Academy became a major centre for the study of mathematics and philosophy; almost all the noteworthy mathematical achievements of the fourth century BC are due to his friends and pupils. After 300 BC, the seat of mathematical learning shifted to Alexandria, but the Academy remained pre-eminent in philosophy until finally it was closed in AD 529 as a result of its support for pagan learning.

Much of what we know about Plato's teaching comes from a series of works in the form of dialogues between Socrates and various historical and fictitious people. As far as astronomy is concerned, two dialogues are particularly important: the *Republic* and the *Timaeus*. We can gauge some of Plato's opinions about astronomy from the following passage in the *Republic* where Socrates says:

> Thus we must pursue astronomy in the same way as geometry, dealing with its fundamental questions. But what is seen in the Heavens must be ignored if we truly want to have our share in astronomy . . . Although celestial phenomena must be regarded as the most beautiful and perfect of that which exists in the visible world (since they are formed of something visible), we must, nevertheless, consider them

> as far inferior to the true, that is to the motions . . . really existing behind them. This can be seen by reason and thought, but not perceived with the eyes.[24]

Plato's ideas on the structure of the Universe are found in his cosmological myth, the *Timaeus*, Timaeus being the main character in the dialogue and through whom Plato expresses Pythagorean ideas. In Plato's universe, the Earth is fixed at the centre and around it are the spheres of the Moon, Sun, Venus, Mercury, Mars, Jupiter, Saturn and the fixed stars.[25] Plato's model explained only the crudest of observed phenomena, but it laid the foundations for the construction of the first mathematical model of the Universe by Plato's collaborator, Eudoxus. The cosmology that Plato described in the *Timaeus* had a great influence on Western thought in the Middle Ages, in part because this was the only one of his works that was available in Latin before the twelfth century.

In the *Timaeus*, Plato described his teachings on the creation of the Universe and the composition of nature. Like most of the physicists of his time, he followed the Sicilian philosopher Empedocles and distinguished four elements of matter: fire, earth, water, and air. The idea that all things are made from these four elements was a synthesis of the various views of the early Ionian philosophers, and may well have been derived from a natural misinterpretation of the action of fire. When something burns, it appears to go from a complex state to a simpler one; thus, it is resolved into its constituent parts. When green wood is burned, for example, the *fire* clearly is visible, the smoke disappears into *air*, *water* boils off from the ends and the ashes clearly are the nature of *earth*.[26]

A member of Plato's Academy who made considerable progress in mathematics and who, indirectly, influenced astronomy, was Theaetetus; the Platonic dialogue bearing his name was a commemorative tribute by Plato to his friend. Among other things, Theaetetus is credited with the original constructions of the octahedron and icosahedron, two of the regular, or Platonic, solids (polyhedra which have the property that all the faces and all the vertices are identical). The other three (the tetrahedron, cube and dodecahedron) were discovered by the Pythagoreans.[27]

One significant fact about the regular solids that led both Plato and, later, Kepler to develop a physical theory around them, is that there are only five

[24] Quoted from Pedersen (1993).
[25] Detailed descriptions of Plato's cosmic scheme can be found in, for example, Knorr (1990), Pecker (2001), pp. 55–70.
[26] This example is taken from Dampier (1965), p. 21.
[27] See Waterhouse (1972), in which the historical problems associated with tracing the early history of these shapes are discussed in some detail.

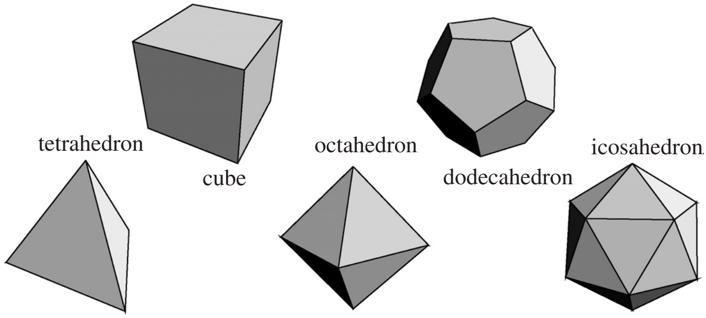

Fig. 1.6. The five Platonic solids.

of them (illustrated in Figure 1.6). This is easy to prove.[28] We use the fact that the sum of the interior angles of the faces which meet at the vertex of a regular polyhedron must be less than 360°. This condition is fulfilled only by three, four, or five equilateral triangles (since the interior angles of an equilateral triangle are 60° and $6 \times 60 = 360$), corresponding to a tetrahedron, octahedron and icosahedron, respectively, three squares (the interior angles of a square are 90° and $4 \times 90 = 360$), corresponding to a cube, and three pentagons (the interior angles of a pentagon are 108° and $4 \times 108 > 360$), corresponding to a dodecahedron. Three hexagons is already too many (the interior angles of a hexagon are 120° and $3 \times 120 = 360$).

Plato was searching constantly for a parallel between the hierarchy of material things and that of mathematical objects, and this led him to consider these solids. The fact that there are only five of them (unlike the regular polygons of which there is an infinite number) made them special. Fire, on account of the shape of a flame, was compared with the tetrahedron, and water, the bulkiest of the elements, corresponded to the icosahedron. Air, having an intermediate density, corresponded to the octahedron, which has a number of triangular faces lying between those of the tetrahedron and icosahedron. Earth was likened to a cube since it is the most immobile of bodies and should be represented by the most stable figure. What about the dodecahedron? Plato got round this by saying that the god had used it for the whole Universe, a statement that did not satisfy many of Plato's disciples; later the dodecahedron became associated with the ether.

As will become evident, many of the ideas about the Universe promoted by people who were the product of Plato's Academy were extremely influential.

[28] The proof given here (which is probably due to Theaetetus) is given in the concluding proposition of Euclid's *Elements*, written in about 300 BC.

But some, which in retrospect we can see to have been correct, were not. One of Plato's pupils, Heraclides, assumed a rotating Earth at the centre of the Universe in order to account for the daily motion of the heavens.

> Heraclides of Pontus and Ecphantus the Pythagorean move the earth, not however in the sense of translation, but in the sense of rotation, like a wheel fixed on an axis, from west to east, about its own centre.[29]

Here, we have an example of a modern view, but one that attracted few followers among Heraclides' contemporaries and which was not considered seriously again until the fourteenth century. This may at first seem strange, but, as with other similar examples, the perceived advantages of Heraclides' theory were far outweighed by the obvious disadvantage of requiring the Earth to spin round, something that is quite alien to everyday experience.[30]

[29] AËTIUS *De placitis* Book III, Volume 13, 3. Translation from Heath (1932). Virtually nothing is known about Ecphantus. Since all his known opinions correspond to those of Heraclides, Heath (1913) thinks it likely that he was in fact a character used by Heraclides in dialogues.

[30] It is often stated that Heraclides believed Mercury and Venus rotated around the Sun rather than the Earth, but Neugebauer (1972) puts this down to a mistranslation of a sentence from Chalcidius' commentary on Plato's *Timaeus.*

.