scend to earth far from the point where his leap began, for the earth would move beneath him while he was in the air. Rocks and trees, cows and men must be hurled from a rotating earth as a stone flies from a rotating sling. Since none of these effects is seen, the earth is at rest. Observation and reason have combined to prove it.

Today in the Western world only children argue this way, and only children believe that the earth is at rest. At an early age the authority of teachers, parents, and texts persuades them that the earth is really a planet and in motion; their common sense is reëducated; and the arguments born from everyday experience lose their force. But reëducation is essential – in its absence these arguments are immensely persuasive – and the pedagogic authorities that we and our children accept were not available to the ancients. The Greeks could only rely on observation and reason, and neither produced evidence for the earth's motion. Without the aid of telescopes or of elaborate mathematical arguments that have no apparent relation to astronomy, no effective evidence for a moving planetary earth can be produced. The observations available to the naked eye fit the two-sphere universe very well (remember the universe of the practical navigator and surveyor), and there is no more natural explanation of them. It is not hard to realize why the ancients believed in the two-sphere universe. The problem is to discover why the conception was given up.

2

THE PROBLEM OF THE PLANETS

Apparent Planetary Motion

If the sun and stars were the only celestial bodies visible to the naked eye, modern man might still accept the fundamental tenets of the two-sphere universe. Certainly he would have accepted them until the invention of the telescope, more than half a century after Copernicus' death. There are, however, other prominent celestial bodies, particularly the planets, and the astronomer's interest in these bodies is the principal source of the Copernican Revolution. Once again we consider observations before dealing with interpretive explanations. And once again the discussion of interpretations will confront us with a new and fundamental problem about the anatomy of scientific belief.

The term planet is derived from a Greek word meaning "wanderer," and it was employed until after Copernicus' lifetime to distinguish those celestial bodies that moved or "wandered" among the stars from those whose relative positions were fixed. For the Greeks and their successors the sun was one of the seven planets. The others were the moon, Mercury, Venus, Mars, Jupiter, and Saturn. The stars and these seven planets were the only bodies recognized as celestial in antiquity. No additional planets were discovered until 1781, long after the Copernican theory had been accepted. Comets, which were well known in the ancient world, were not considered celestial bodies before the Copernican Revolution (Chapter 6).

All of the planets behave somewhat like the sun, though their motions are uniformly more complex. All have a westward diurnal motion with the stars, and all move gradually eastward among the stars until they return to approximately their original positions. Throughout their motions the planets stay near the ecliptic, sometimes wandering north of it, sometimes south, but very seldom leaving the band of the zodiac, an imaginary strip in the sky extending for 8° on either side of the ecliptic. At this point the resemblance between planets ends, and the study of planetary irregularities begins.

The moon travels around the ecliptic faster and less steadily than the sun. On the average it completes one journey through the zodiac in 27⅓ days, but the time required for any single journey may differ from the average by as much as 7 hours. In addition, the appearance of the moon's disk changes markedly as it moves. At new moon its disk is completely invisible or very dim; then a thin bright crescent appears, which gradually waxes until, about a week after new moon, a semicircular sector is visible. About 2 weeks after new moon the full circular disk appears; then the cycle of phases is reversed, and the moon gradually wanes, reaching new moon again about 1 month after the preceding new moon. The cycle of phases is recurrent, like the moon's journey through the signs of the zodiac, but the two lunar cycles are significantly out of step. New moon recurs after an average interval of 29½ days (individual cycles may differ by as much as ½ day from this average), and, since this is 2 days longer than the period of an average journey around the zodiac, the position of successive new moons must gradually move eastward through the constellations. If new moon occurs at the position of the vernal equinox one month, the moon will still be waning when it returns to the vernal equinox 27⅓ days later. New moon does not recur for about 2 days more, by which time the moon has moved almost 30° east from the equinox.

Because they are easily visible and conveniently spaced, the moon's phases provided the oldest of all calendar units. Primitive forms of both the week and the month appear in a Babylonian calendar from the third millennium B.C., a calendar in which each month began with the first appearance of the crescent moon and was subdivided at the 7th, 14th, and 21st days by the recurrent "quarters" of the moon's cycle. At the dawn of civilization men must have counted new moons and quarters to measure time intervals, and as civilization progressed they repeatedly attempted to organize these fundamental units into a coherent long-term calendar – one that would permit the compilation of historical records and the preparation of contracts to be honored at a specified future date.

But at this point the simple obvious lunar unit proved intractable. Successive new moons may be separated by intervals of either 29 or 30 days, and only a complex mathematical theory, demanding generations of systematic observation and study, can determine the length of a specified future month. Other difficulties derive from the incommensurable lengths of the average lunar and solar cycles. Most societies (but not all, for pure lunar calendars are still used in parts of the Middle East) must adjust their calendars to the sun-governed annual climatic variation, and for this purpose some systematic method for inserting an occasional thirteenth month into a basic year of 12 lunar months (354 days) must be devised. These seem to have been the first difficult technical problems encountered by ancient astronomy. More than any others, they are responsible for the birth of quantitative planetary observation and theory. The Babylonian astronomers who finally solved these difficulties between the eighth and third centuries B.C., a period during most of which Greek science was still in its infancy, accumulated much of the fundamental data subsequently incorporated into the developed structure of the two-sphere universe.

Unlike the moon and sun, the remaining five planets appear as mere points of light in the heavens. The untrained naked-eye observer can distinguish them from stars with assurance only by a series of observations that discloses their gradual motion around the ecliptic. Usually the planets move eastward through the constellations: this is their so-called "normal motion." On the average, both Mercury and Venus require 1 year for each complete circuit of the zodiac; the length of Mars's cycle averages 687 days; Jupiter's average period is 12 years; and Saturn's is 29 years. But in all cases the time required for any single journey may be quite different from the average period. Even when moving eastward through the stars, a planet does not continue at a uniform rate.

Nor is its motion uniformly eastward. The normal motion of all planets except the sun and moon is occasionally interrupted by brief intervals of westward or "retrograde" motion. Compare Mars retrogressing in the constellation Taurus, shown in Figure 15, with the normal motion of the sun through Taurus, shown in Figure 9 (p. 22). Mars enters the diagram in normal (eastward) motion, but as its motion continues, the planet gradually slows until at last it reverses its direction and begins to move westward, in retrograde. Other planets behave in much the same way, each one repeating the interlude of retrograde motion after a fixed length of time. Mercury briefly reverses its motion through the stars once every 116 days, and Venus retrogresses every 584 days. Mars, Jupiter, and Saturn show retrograde motion every 780, 399, and 378 days, respectively.

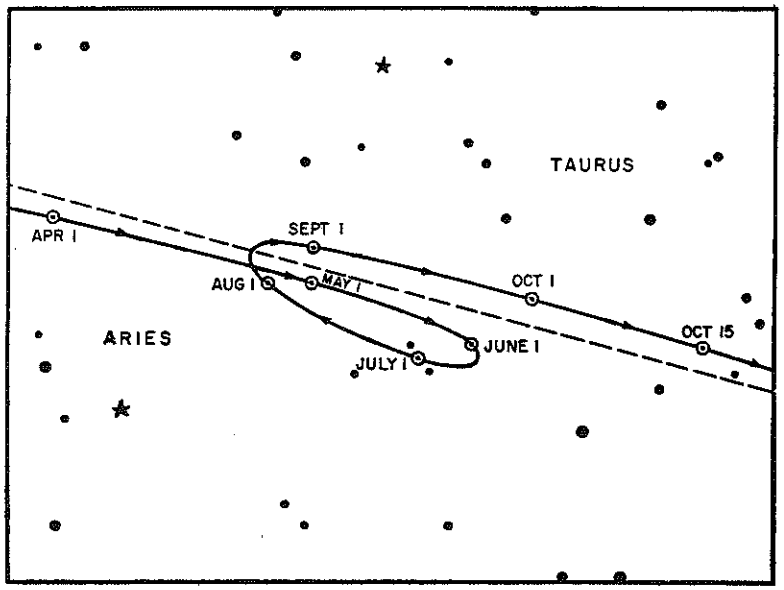

Figure 15. Mars retrogressing in Aries and Taurus. The section of sky is the same as that shown in Figure 9 and in the box on the star map of Figure 8. The broken line is the ecliptic and the solid line the path of the planet. Note that Mars does not stay on the ecliptic and that, though its over-all motion is eastward among the stars, there is a period from the middle of June to early August during which it moves to the west. The retrogressions of Mars are always of approximately this form and duration, but they do not always occur on the same date or in the same part of the sky.

In their gradual eastward motions interrupted by periodic westward retrogressions, the five wandering stars behave quite similarly. But there is an additional characteristic of their motion which divides them into two groups; this is the correlation between their position and the sun's. Mercury and Venus, the two so-called inferior planets, never get very far from the sun. Mercury is always found within 28° of the sun's moving disk, and Venus's maximum "elongation" is 45°. Both planets move in a continuous slow shuttle, back and forth across the

moving sun; for a time they move eastward with the sun, then retrogress across its disk, and finally reverse themselves to overtake the sun once more. When to the east of the sun, either of these inferior planets appears as an "evening star," becoming visible shortly after sunset and then rapidly following the sun below the horizon. After retrogressing westward across the sun's disk, the planet becomes a "morning star," rising shortly before dawn and disappearing in the brilliant light of sunrise. But in between, when close to the sun, neither Mercury nor Venus can be seen at all. Therefore, until their motion was analyzed with respect to the sphere of the stars, neither of the inferior planets was recognized as the same celestial body when it appeared as a morning and as an evening star. For millenniums Venus had one name when it rose in the east shortly before dawn and another when, weeks later, it again became visible just over the western horizon shortly after sunset.

Unlike Mercury and Venus, the superior planets, Mars, Jupiter, and Saturn, are not restricted to the same part of the sky as the sun. Sometimes they are very close to or "in conjunction" with it; at other times they are 180° across the sky or "in opposition" to the sun; between these times they assume all the intermediate positions. But though their positions are unrestricted, their behavior does depend upon their relation to the sun. Superior planets retrogress only when they are in opposition. Also, when in retrograde motion across the sky from the sun, superior planets appear brighter than at any other time. This increased brilliance, which has usually been interpreted (at least since the fourth century B.C.) as indicating a decrease in the planet's distance from the earth, is particularly striking in the case of Mars. Normally a relatively inconspicuous planet, Mars in opposition will frequently outshine every celestial body in the night sky except the moon and Venus.

Interest in the five wandering stars is by no means so ancient as a concern with the sun and moon, presumably because the wandering stars had no obvious practical bearing upon the lives of ancient peoples. Yet observations of the appearance and disappearance of Venus were recorded in Mesopotamia as early as 1900 B.C., probably as omens, portents of the future, like the signs to be read in the entrails of sacrificial sheep. These scattered observations presage the much later development of systematic astrology, a means of forecasting whose intimate relation to the development of planetary astronomy is considered in the next chapter. The same concern with omens clearly motivated the more systematic and complete records of eclipses, retrograde motions, and other striking planetary phenomena compiled by Babylonian observers after the middle of the eighth century B.C. Ptolemy, the dean of ancient astronomers, later complained that even these records were fragmentary, but fragmentary or not they provided the first data capable of specifying the full-scale problem of the planets as that problem was to develop in Greece after the fourth century B.C.

The problem of the planets is partially specified by the description of the planetary motions sketched in the preceding pages. How are the complex and variable planetary motions to be reduced to a simple and recurrent order? Why do the planets retrogress, and how account for the irregular rate of even their normal motions? These questions indicate the direction of most astronomical research during the two millenniums from the time of Plato to the time of Copernicus. But because it is almost entirely qualitative, the preceding description of the planets does not specify the problem fully. It states a simplified problem and in some respects a misleading one. As we shall shortly see, qualitatively adequate planetary theories are easily invented: the description above can be reduced to order in several ways. The astronomer's problem, on the other hand, is by no means simple. He must explain not merely the existence of an intermittent westward motion superimposed upon an over-all eastward motion through the stars, but also the precise position that each planet occupies among the stars on different days, months, and years over a long period of time. The real problem of the planets, the one that leads at last to the Copernican Revolution, is the quantitative problem described in lengthy tables which specify in degrees and minutes of arc the varying position of every planet.

The Location of the Planets

The two-sphere universe, as developed in the last chapter, provided no explicit information about the positions or motions of the seven planets. Even the sun's location was not discussed. To appear "at" the vernal equinox (or any other point on the stellar sphere) the sun need merely be somewhere on a line stretching from the observer's eye to or through the appropriate point in the background of stars.

Like the other planets, it might be either inside, on, or perhaps even outside the sphere of the stars. But though the two-sphere universe fails to specify the shape or location of the planetary orbits, it does make certain choices of position and orbit more plausible than others, and it therefore at once guides and restricts the astronomer's approach to the problem of the planets. That problem was set by the results of observation, but, from the fourth century B.C., it was pursued in the conceptual climate of two-sphere cosmology. Both observation and theory made essential contributions to it.

Within a two-sphere cosmology, for example, the planetary orbits should if possible preserve and extend the fundamental symmetry embodied in the first two spheres. Ideally the orbits should therefore be earth-centered circles, and the planets should revolve in these circles with the same regularity that is exemplified in the rotation of the stellar sphere. The ideal does not quite conform to observation. As we shall see presently, an earth-centered circular orbit located in the plane of the ecliptic provides a good account of the sun's annual motion, and a similar circle can give an approximate account of the somewhat less regular motion of the moon. But circular orbits do not even hint at an explanation of the gross irregularities, like retrogression, observed in the motions of the other five wandering "stars." Nevertheless, astronomers who believed in the two-sphere universe could, and for centuries did, think that earth-centered circles were the natural orbits for planets. Such orbits at least accounted for the over-all average eastward motions. Observed deviations from the average motion – changes in the rate or direction of a planet's motion – indicated that the planet itself had deviated from its natural circular orbit, to which it would again return. On this analysis the problem of the planets became simply that of explaining the observed deviation from average motion through the stars in terms of a corresponding deviation of each planet from its single circular orbit.

We shall examine some of the ancient explanations of these deviations in the next three sections, but first notice, as the ancients also did, how far it is possible to proceed by neglecting the planetary irregularities and assuming simply that all orbits are at least approximately circular. Almost certainly, in the two-sphere universe, the planets move in the region between the earth and the stars. The stellar sphere itself was often viewed as the outer boundary of the universe, so that the planets could not be outside it; the difference between planetary and stellar motions indicated that the planets were probably not located on the sphere, but in some intermediate region where they were affected by some influence that was inactive at the stellar sphere; the whole argument gained force from the detail visible on the face of the moon, presumptive evidence that one planet, at least, must be nearer than the stars. Ancient astronomers, therefore, laid out the planetary orbits in that vast and previously empty space between the earth and the sphere of the stars. By the end of the fourth century B.C., the two-sphere universe was filling up. Later it was to become crowded.

Once the general location and shape of their orbits were known, it proved possible to make a plausible and satisfying guess about the order in which the planets were arranged. Planets like Saturn and Jupiter, whose eastward motion was slow and whose total motion, therefore, very nearly kept pace with the stars, were supposed to be close to the stellar sphere and far from the earth. The moon, on the other hand, which loses over 12° a day in its race with the stars, must be closer to the stationary surface of the earth. Some ancient philosophers seem to have justified this hypothetical arrangement by imagining that the planets floated in a gigantic aethereal vortex, whose outer surface moved rapidly with the sphere of the stars and whose interior was at rest at the earth's surface. Any planet caught in such a vortex would lose more ground with respect to the stellar sphere if it were closer to the earth. Other philosophers reached the same conclusion by a different sort of argument, later recorded, at least in its essentials, by the Roman architect Vitruvius (first century B.C.). In analyzing the differences between the intervals required by different planets for trips about the ecliptic, Vitruvius suggested an illuminating analogy:

> Place seven ants on a wheel such as potters use, having made seven channels on the wheel about the center, increasing successively in circumference; and suppose those ants obliged to make a circuit in these channels while the wheel is turned in the opposite direction. In spite of having to move in a direction contrary to that of the wheel, the ants must necessarily complete their journeys in the opposite direction, and that ant which is nearest the center must finish its circuit sooner, while the ant that is going round at the outer edge of the disk of the wheel must, on account of the size of its circuit, be much slower in completing its course, even though it is moving just as quickly as the other. In the same way, these stars, which struggle on against the course of the firmament, are accomplishing an orbit on paths of their own; but, owing to the revolution of the heaven, they are swept back as it goes round every day.[1]

Before the end of the fourth century B.C., arguments like the above had led to an image of the universe similar to the one sketched in Figure 16; diagrams like this, or their verbal equivalents, remained current in elementary books on astronomy or cosmology until the early seventeenth century, long after Copernicus' death. The earth is at the center of the stellar sphere, which bounds the universe; immediately inside this outer sphere is the orbit of Saturn, the planet that takes longest to move around the zodiac; next comes Jupiter and then Mars.

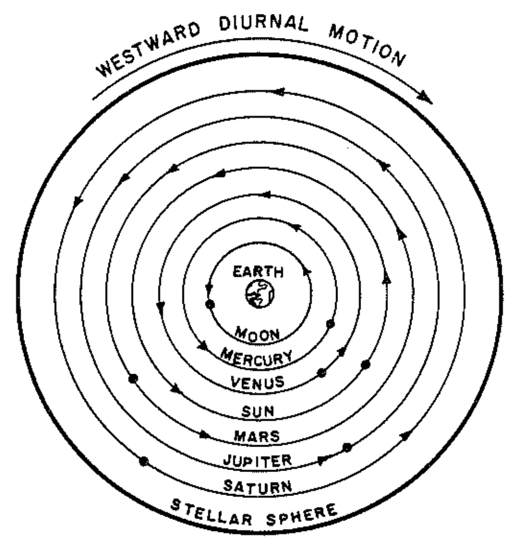

Figure 16. Approximate planetary orbits in the two-sphere universe. The outermost circle is a cross section of the stellar sphere in the plane of the ecliptic.

To this point the order is unambiguous: the planets are arranged, from the outside, in the order of decreasing orbital period; the same technique places the lunar orbit closest to the earth. But the remaining three planets present a problem; the sun, Venus, and Mercury all complete their journeys about the earth in the same average time, 1 year, and their order therefore cannot be determined by the device applied to the other planets. There was, in fact, much disagreement about their order during antiquity. Until the second century B.C. most astronomers placed the sun's orbit just outside the moon's, Venus's outside the sun's, then Mercury's, and then Mars's. After that date, however, the order shown in the diagram – moon, Mercury, Venus, sun, Mars, etc. – became increasingly popular. In particular, it was adopted by Ptolemy, and his authority imposed it upon most of his successors. We shall therefore take this order as standard throughout the early chapters of this book.

As a structural diagram Figure 16 is still very crude. It gives no meaningful indication of relative dimensions of the various orbits, and it makes no attempt to provide for the observed planetary irregularities. But the conception of the universe embodied in the diagram had two important functions in the subsequent development of astronomy and cosmology. In the first place, the diagram contains most of the structural information about the earth-centered universe that ever became common knowledge among nonastronomers. The further achievements of ancient astronomy, to which we shall turn in a moment, were too mathematical for most laymen to understand. As the next two chapters indicate more fully, the most influential cosmologies developed in antiquity and the Middle Ages did not follow ancient astronomy very far beyond this point. Astronomy now becomes esoteric; its further development does not provide man with a home.

In addition, the structural diagram in Figure 16 is, despite its crudity, an immensely powerful tool in astronomical research. For many purposes it proved both economical and fruitful. For example, during the fourth century B.C., the concepts embodied in the diagram provided a complete qualitative explanation of both the phases of the moon and lunar eclipses; during the fourth and third centuries these same concepts led to a series of relatively accurate determinations of the circumference of the earth; and during the second century B.C., they provided the basis for a brilliantly conceived estimate of the sizes and distances of the sun and moon. These explanations and measurements, particularly the last, typify the immense ingenuity and power of the ancient astronomical tradition. They are, however, here relegated to the Technical Appendix (Sections 3 and 4), because they were not affected by the change in astronomical theory during the Copernican Revolution. Nevertheless, they are relevant to the Revolution. The ability of the developed two-sphere universe to explain and ultimately to predict prominent celestial phenomena like eclipses, as well as its

ability to specify some linear dimensions of the celestial regions, immeasurably increased the hold of the two-sphere conceptual scheme upon the minds of both astronomers and laymen.

These achievements do not, however, touch upon the fundamental problem posed by the continuous irregularity of planetary motion, and this problem provides the pivot upon which the Copernican Revolution ultimately turns. Like so many other problems of ancient astronomy, it seems first to have emerged during the fourth century B.C., when the two-sphere universe, by explaining diurnal motion, enabled Greek astronomers to isolate the residual planetary irregularities for the first time. During the following five centuries successive attempts to explain these irregularities produced several planetary theories of unprecedented accuracy and power. But these attempts also constitute the most abstruse and mathematical part of ancient astronomy, and they are therefore usually omitted from books like this. Though a simplified précis of ancient planetary theory seems a minimal requisite for an understanding of the Copernican Revolution, a few readers may prefer to skim the next three sections (particularly the first, in which the technical presentation is especially compact), picking up the narrative again with the discussion of scientific belief that closes this chapter.

The Theory of Homocentric Spheres

The philosopher Plato, whose searching questions dominated so much of subsequent Greek thought, seems to have been the first to enunciate the problem of the planets, too. Early in the fourth century B.C. Plato is said to have asked: "What are the uniform and ordered movements by the assumption of which the apparent movements of the planets can be accounted for?" [2] and the first answer to this question was provided by his onetime pupil Eudoxus (*c.*408 – *c.*355 B.C.). In Eudoxus' planetary system each planet was placed upon the inner sphere of a group of two or more interconnected, concentric spheres whose simultaneous rotation about different axes produced the observed motion of the planet. Figure 17*a* shows a cross section of two such interlocked spheres whose common center is the earth and whose points of contact are the ends of the slanted axis of the inner sphere, which serve as pivots. The outer sphere is the sphere of the stars, or at least it has the same motion as that sphere; its axis passes through the north and south celestial poles, and it rotates westward about this axis once in 23 hours 56 minutes. The inner sphere's axis makes contact with the outer sphere at two diametrically opposite points 23½° away from the north and south celestial poles; therefore the equator of the inner sphere, viewed from the earth, always falls on the ecliptic of the sphere of the stars, regardless of the rotation of the two spheres.

If the sun is now placed at a point on the equator of the inner sphere and that sphere is turned slowly eastward about its axis once in a year while the outer sphere turns about its axis once a day, the sum of the two motions will reproduce the observed motion of the sun. The outer sphere produces the observed westward diurnal motion of rising and setting; the inner sphere produces the slower annual eastward motion of the sun, around the ecliptic. Similarly, if one eastward rotation of the inner sphere occurs every 27⅓ days and if the moon is placed on the equator of this sphere, then the motion of this inner sphere must produce the average motion of the moon around the ecliptic. The north and south deviations of the moon from the ecliptic and some of the irregularities in the time required by the moon for successive journeys can be approximated by adding one more very slowly moving sphere to the system. Eudoxus also used (though unnecessarily) a third sphere to describe the motion of the sun, so that six spheres were required to treat the moon and sun together.

The spheres shown in Figure 17*a* were known as homocentric spheres, because they have a common center, the earth. Two or three such spheres can approximately represent the total motion of the sun and of the moon, but they cannot account for the retrograde motions of the planets, and Eudoxus' greatest genius as a geometer was displayed in the modification of the system that he introduced in treating the apparent behavior of the remaining five planets. For each of these he used a total of four spheres, sketched in cross section in Figure 17*b*. The two outer spheres move just like the spheres of Figure 17*a*: the outer sphere has the diurnal motion of the sphere of the stars and the second sphere (counting from the outside) turns eastward once in the *average* time required by the planet to complete a journey around the ecliptic. (Jupiter's second sphere, for example, turns once in 12 years.) The third sphere is in contact with the second sphere at two diametrically opposite points on the ecliptic (the equator of the second sphere), and the axis of the fourth or innermost sphere is fastened to the third sphere at an angle that depends upon the

characteristics of the motion to be described. The planet itself (Jupiter in the example above) is located on the equator of the fourth sphere.

Suppose now that the two outer spheres are held stationary and that the two innermost spheres rotate in opposite directions, each completing one axial rotation in the interval that separates two suc-

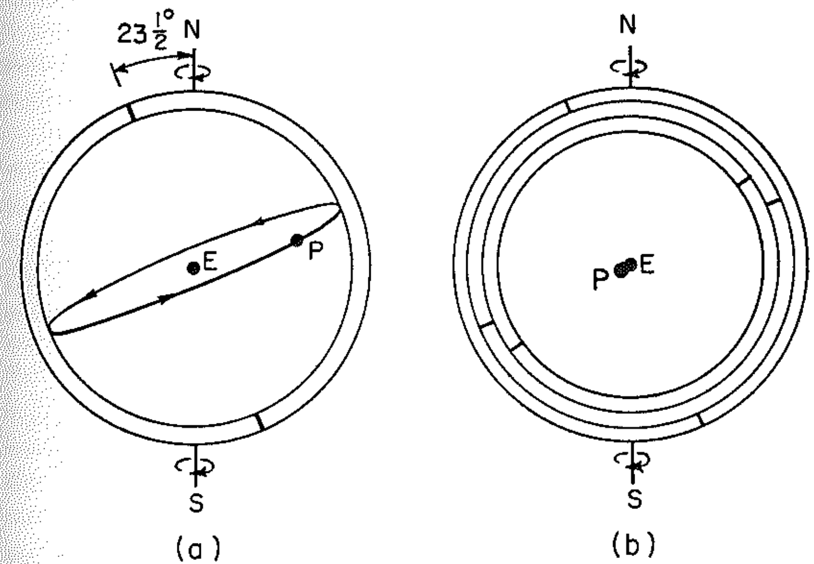

Figure 17. Homocentric spheres. In the two-sphere system (*a*) the outermost sphere produces the diurnal rotation, and the inner sphere moves the planet (sun or moon) steadily eastward around the ecliptic. In the four-sphere system (*b*) the planet *P* lies out of the plane of the paper, almost on a line from the earth *E* to the reader's eye. The two innermost spheres then produce the looped motion shown in Figure 18, and the two outer spheres produce both the diurnal motion and the average eastward planetary drift.

cessive retrogressions of the planet (399 days for Jupiter). An observer watching the motion of the planet against the temporarily stationary second sphere will see it move slowly in a figure eight, both of whose loops are bisected by the ecliptic. This is the motion sketched in Figure 18; the planet passes slowly around the loops from positions 1 to 2, 2 to 3, 3 to 4, . . . , spending equal times between each numbered point and the next, and returning to its starting point after the interval between retrogressions. During its motion from 1 to 3 to 5 the planet moves eastward along the ecliptic; during the other half of the time, while the planet moves from 5 to 7 and back to 1, it moves westward.

Now allow the second sphere to rotate eastward, carrying the two rotating inner spheres with it, and suppose that the total motion of the planet is observed against the background of stars on the first sphere, again held temporarily stationary. At all times the planet is moved eastward by the motion of the second sphere, and half of the time (while it moves from 1 to 5 in Figure 18) the planet receives an additional eastward motion from the two inner spheres so that its net motion is eastward and even more rapid than that of the second sphere alone. But during the other half of the time (as the planet moves from 5 to 1 in Figure 18) the eastward motion of the second sphere is opposed by a westward motion due to the two inner spheres, and, when this westward motion is most rapid (near 7 in Figure 18), the net motion of the planet against the sphere of the stars may actually be to the west, in the retrograde direction. This is just the characteristic of observed planetary motions that Eudoxus was striving to reproduce in his model.

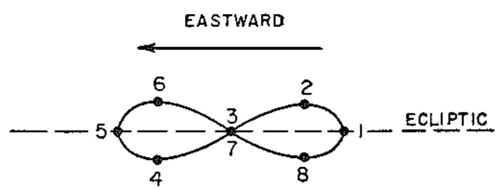

Figure 18. The looped motion generated by the two innermost homocentric spheres. In the full four-sphere system this looped motion is combined with the steady eastward motion of the second sphere, a motion that by itself would carry the planet around the ecliptic at a uniform rate. When the looped motion is added, the total motion of the planet varies in rate and is no longer confined to the ecliptic. While the planet moves from 1 to 5 on the loop, its total motion is more rapid than the average eastward motion generated by the second sphere; while the planet moves back from 5 to 1 on the loop, its eastward motion is slower than that produced by the second sphere, and when it gets near 3, it may actually move westward, in retrograde.

A system of four interlocked homocentric spheres approximates the retrograde motion of Jupiter, and a second set of four spheres can account for the motion of Saturn. For each of the three remaining planets, five spheres are needed (an extension provided by Eudoxus' successor Callippus around 330 B.C.), and analysis of the resulting motions becomes correspondingly more complex. Fortunately, we need not pursue these complex combinations of rotating spheres further, because all homocentric systems have one severe drawback which in antiquity led to their early demise. Since Eudoxus' theory places each planet on a sphere concentric with the earth, the distance between a planet and the earth cannot vary. But planets appear brighter, and therefore seem closer to the earth, when they retrogress. During antiquity the homocentric system was frequently criticized for its failure to explain this variation in planetary brilliance, and the system was abandoned by most astronomers almost as soon as a more adequate explanation of the appearances was proposed.

But though short-lived as a significant astronomical device, homocentric spheres play a major role in the development of astronomical and cosmological thought. By a historical accident the century during which they seemed to provide the most promising explanation of planetary motion embraced most of the lifetime of the Greek philosopher Aristotle, who incorporated them in the most comprehensive, detailed, and influential cosmology developed in the ancient world. No comparably complete cosmology ever incorporated the mathematical system of epicycles and deferents which, in the centuries after Aristotle's death, was employed to explain planetary motion. The conception that planets are set in rotating spherical shells concentric with the earth remained an accepted portion of cosmological thought until early in the seventeenth century. Even the writings of Copernicus show important vestiges of this conception. In the title of Copernicus' great work, *De Revolutionibus Orbium Caelestium*, the "orbs" or spheres are not the planets themselves but rather the concentric spherical shells in which the planets and the stars are set.

Epicycles and Deferents

The origin of the device that replaced homocentric spheres in explaining the details of planetary motion is unknown, but its features were early investigated and developed by two Greek astronomers and mathematicians, Apollonius and Hipparchus, whose work spans the period from the middle of the third century to the end of the second century B.C. In its simplest form (Figure 19*a*) the new mathematical mechanism for the planets consists of a small circle, the epicycle, which rotates uniformly about a point on the circumference of a second rotating circle, the deferent. The planet, *P*, is located on the epicycle, and the center of the deferent coincides with the center of the earth.

The epicycle-deferent system is intended to explain only motion with respect to the sphere of the stars. Both the epicycle and the deferent in Figure 19*a* are drawn on the plane of the ecliptic, so that the rotation of the stellar sphere carries the entire diagram (except the central earth) through one rotation per day and thus produces the diurnal motion of the planet. If the epicycle and deferent of the figure were stationary and did not have an additional motion of their own, the planet would be fixed in the plane of the ecliptic and would therefore have the motion of a zodiacal star, a westward circle executed once in every 23 hours 56 minutes. From now on, whenever reference is made to the motion of the deferent or the epicycle, it is the *additional* motion of these circles in the plane of the ecliptic that is meant. The diurnal rotation of the sphere and of the plane of the ecliptic will be taken for granted.

Suppose, for example, that the deferent rotates eastward once in a year and that the sun is placed on the deferent at the position now occupied by the center of the epicycle, the epicycle itself being removed. Then the rotation of the deferent carries the sun through its annual journey around the ecliptic, and the sun's motion has been analyzed, at least approximately, in terms of the motion of a single deferent in the plane of the ecliptic. This is the technique taken for granted in the explanation of average planetary motions in Figure 16.

Now imagine that the sun is removed and the epicycle is returned to its position on the deferent. If the epicycle rotates just three times around its moving center while the deferent rotates once and if the two circles rotate in the same direction, then the total motion of the planet within the sphere of the stars produced by the combined motions of the epicycle and the deferent is just the looped curve shown in Figure 19*b*. When the rotation of the epicycle carries the planet outside of the deferent, the motions of both the epicycle and the deferent combine to move the planet to the east. But when the motion of the epicycle places the planet well inside the deferent, the epicycle carries the planet westward, in opposition to the motion of the deferent. Therefore, when the planet is closest to the earth, the two motions may combine to produce a net westward or retrograde motion. In Figure 19*b* the planet retrogresses whenever it is on the interior part of one of the small loops; everywhere else the planet moves normally toward the east, but at a variable rate.

Figure 19*c* shows the motion of the planet through one of the loops as viewed against the sphere of the stars by an observer on earth. Since both observer and loop are in the same plane, that of the ecliptic, the observer cannot see the open loop itself. What he sees is merely the position of the planet against the background provided by the ecliptic. Thus as the planet moves from position 1 to 2 in Figures 19*b* and 19*c*, the observer sees it move along the ecliptic toward the east. As the planet approaches position 2, it appears to move more slowly, stopping momentarily at 2 and then moving westward along the ecliptic as it travels from 2 towards 3. Finally the westward journey of

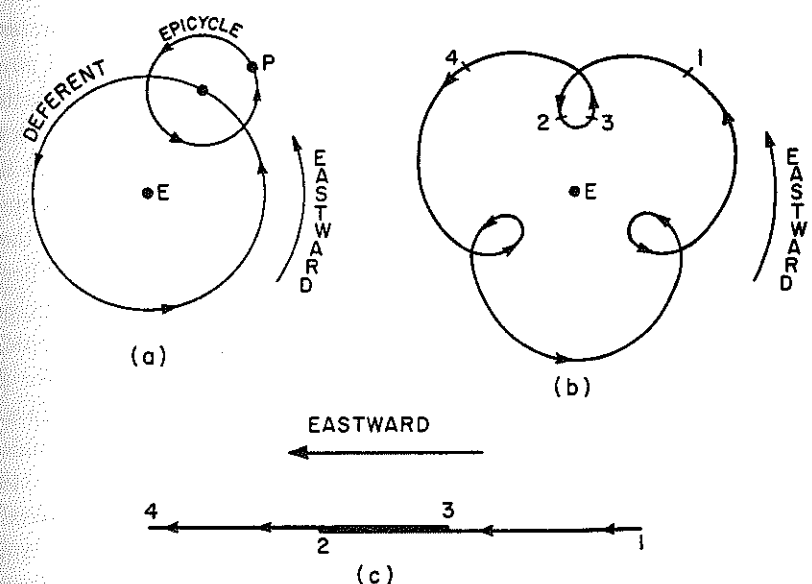

Figure 19. The basic epicycle-deferent system. A typical deferent and epicycle are shown in (*a*); the looped motion that they generate in the plane of the ecliptic is illustrated in (*b*); the third diagram (*c*) shows a portion (1–2–3–4) of the motion in (*b*) as it is seen by an observer on the central earth, *E*.

the planet on the ecliptic halts, and the planet again moves eastward, leaving position 3 on the loop and moving toward 4.

A system of one epicycle and one deferent therefore carries a planet around the ecliptic in an interval that, on the average, exactly equals the time required for one revolution of the deferent. The eastward motion is, however, interrupted and the planet temporarily moves westward at regular intervals equal to the time required for one revolution of the epicycle. The rates of revolution of the epicycle and deferent may be adjusted to fit the observations for any planet, yielding just that intermittent eastward motion among the stars which planets are observed to have. Furthermore, the epicycle-deferent system reproduces one other important qualitative feature of the appearances: a planet can retrogress only when its motion brings it nearest to the earth and that is the position in which the planet should and does appear brightest. Its great simplicity plus this novel explanation of varying planetary brilliance are the primary causes for the new system's victory over the older system of homocentric spheres.

The epicycle-deferent system described by Figure 19 incorporates one special simplification that is not characteristic of the motion of any planet. The epicycle is made to revolve *exactly* three times for each revolution of the deferent. Therefore, whenever the deferent completes one revolution, the epicycle returns the planet to the same position it occupied at the beginning of the revolution; the retrograde loops always occur at the same places; and the planet always requires the same amount of time to complete its trip around the ecliptic. When designed to fit the observations of real planets, however, epicycle-deferent systems never perform in quite this manner. For example, Mercury is observed to require an average of 1 year to complete a journey around the ecliptic, and it retrogresses once every 116 days. Therefore Mercury's epicycle must revolve just over three times while the deferent turns once; the epicycle completes three revolutions in 348 days which is less than the year required for a rotation of the deferent.

Figure 20*a* shows the path of a planet carried through one trip around the ecliptic by an epicycle that turns slightly more than three times for each rotation of its deferent. The planet starts in the middle of a retrograde loop and completes its third full loop before the deferent completes its first full rotation; the planet therefore averages slightly more than three retrograde loops in each trip around the ecliptic. If the motion of Figure 20*a* were continued through a second trip, the new set of retrograde loops would fall slightly to the west of those generated during the first trip. Retrograde motion would not occur at the same position in the zodiac on successive trips, and this is characteristic of the observed progress of planets along the ecliptic.

Figure 20*b* indicates a second characteristic of the motion generated by an epicycle that does not revolve an integral number of times in each revolution of the deferent. The planet at *P* in the figure is at the position closest to the earth, the position from which the journey of Figure 20*a* began. After one revolution of the deferent, the epicycle will have turned slightly more than three times, and the planet will have arrived at position *P′*, so that it now appears to the west of its

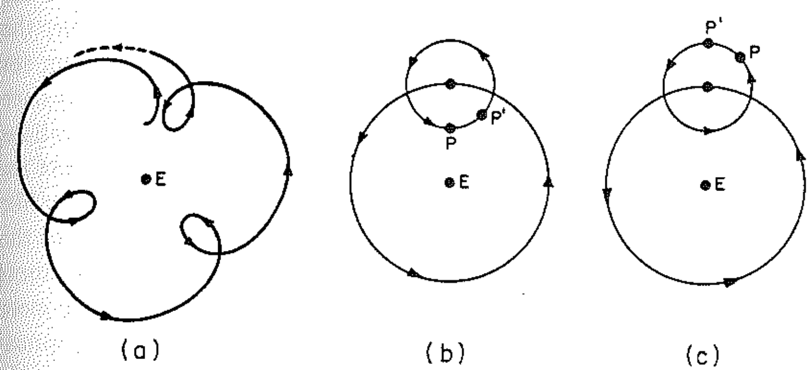

Figure 20. Motion generated by an epicycle and deferent when the epicycle turns slightly more than three times for each revolution of the deferent. The planet's path during a single complete journey through the stars is shown in (*a*). This journey requires more than one revolution of the deferent, as indicated by (*b*), which shows the planet's position at the beginning (*P*) and the end (*P′*) of the deferent's first full revolution. Diagram (*c*) shows the planet's position at the beginning and end of a later revolution of the deferent, one that carries the planet more than once around the ecliptic.

starting point. The deferent must turn eastward through more than a single revolution to carry the planet fully around the ecliptic; the corresponding trip through the constellations therefore requires more time than the average. Others, however, require less. After several more revolutions of the deferent, each ending with the planet still farther from the earth, the planet might start a new journey from the new position *P* in Figure 20*c*. One more revolution of the deferent would carry the planet to *P′*, a point to the east of *P* Since this revolution of the deferent carries the planet through more than one trip around the ecliptic, this journey is a particularly rapid one. Figures 20*b* and 20*c* represent very nearly the extreme values of the time required for a journey around the ecliptic; intermediate trips consume intermediate amounts of time; on the average, a journey around the ecliptic requires the same time as a rotation of the deferent. But the epicycle-deferent system allows for deviations from one trip to the next. Once again it provides an economical explanation of an observed irregularity of the planetary motions.

To describe the motions of all the planets a separate epicycle-deferent system must be designed for each. The motion of the sun and moon can be treated approximately by a deferent alone, for these planets do not retrogress. The sun's deferent turns once a year; the moon's revolves once in 27⅓ days. The epicycle-deferent system for Mercury is much like the one discussed above; the deferent turns once a year and the epicycle once in 116 days. By utilizing the observations recorded early in this chapter, we could design similar systems for other planets. Most of these would yield looped planetary paths like the one shown in Figure 20*a*. If the epicycle is larger relative to the deferent, the size of the loops is increased. If the epicycle turns more quickly relative to the speed of the deferent, then there are more loops included in one journey around the ecliptic. There are approximately eleven loops in each trip made by Jupiter, and approximately twenty-eight in each by Saturn. In short, by appropriate variations in the relative sizes and the speeds of the epicycle and the deferent, this system of compounded circular motions can be adjusted to fit approximately an immense variety of planetary motions. A properly designed combination of circles will even give a good qualitative account of the immense irregularities in the motion of an atypical planet like Venus (Figure 21).

Ptolemaic Astronomy

The discussion of the previous section illustrates the power and versatility of the epicycle-deferent system as a method for ordering and predicting the motions of the planets. But this is only the first step. Once the system was available to account for the most striking irregularities of planetary motion – retrogression and the irregular amounts of time consumed in successive journeys around the ecliptic – it became clear that there were still other, though very much smaller, irregularities to be considered.

Just as the two-sphere model provided a precise mechanism for the diurnal motions, thus permitting detailed study of the principal planetary irregularities, so the epicycle-deferent system, by providing

an account of the main planetary motions, permitted the observational isolation of smaller irregularities. This is the first example of the con-

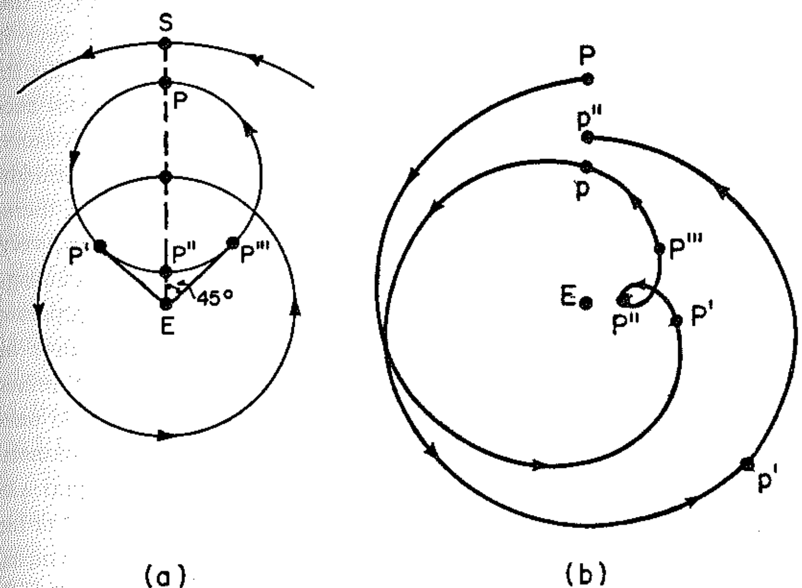

Figure 21. (*a*) A one-epicycle one-deferent system for Venus and (*b*) the motion that it generates in the plane of the ecliptic.

In (*a*), notice the following characteristics of the design: The deferent rotates once in a year, so that if the center of the epicycle is once aligned with the earth, *E*, and the center of the sun, *S*, it will stay in alignment forever, and Venus will never appear very far from the sun. The angles *SEP'* and *SEP'''* are the largest angles that can appear between the sun and Venus, and the condition that these angles of maximum elongation be 45° completely determines the relative sizes of the epicycle and deferent. The epicycle rotates once in 584 days, so that if Venus starts at *P*, close to the sun, it will arrive at *P'* (maximum elongation as an evening star) after 219 days (3/8 revolution); at *P''* after 292 days (1/2 revolution); and at *P'''* (maximum elongation as a morning star) after 365 days (5/8 revolution).

The second diagram shows the path along which Venus is carried by the moving circles sketched in (*a*). Here *P* is the starting point, as in the first diagram; *P'* is Venus's position at maximum eastward elongation (219 days); *P''* is the planet's location midway through a retrograde loop (292 days); and *P'''* is its position at maximum westward elongation (365 days). Venus's first journey around the ecliptic ends at *p* after 406 days (note the great length) and includes one retrogression and two maximum elongations. Its next trip (*p* to *p'* to *p''*) requires only 295 days and includes none of these characteristic phenomena. At *p'* Venus is again closest to the sun, a position reached after one complete revolution of the epicycle (584 days). This is, at least qualitatively, the way Venus does behave!

cept's fruitfulness. When the motion predicted by a one-epicycle one-deferent system is compared with the observed motion of an individual planet, it turns out that the planet is not always seen at quite the position on the ecliptic where the geometry of the model says it should be. Venus does not, if observed precisely, always attain its maximum deviation of 45° from the sun; the intervals between successive retrogressions of a single planet are not always quite the same; and none of the planets, except the sun, stays on the ecliptic throughout its motion. The one-epicycle one-deferent system was not, therefore, the final answer to the problem of the planets. It was only a very promising start and one that lent itself to both immediate and long-continued development. During the seventeen centuries that separate Hipparchus from Copernicus all the most creative practitioners of technical astronomy endeavored to invent some new set of minor geometric modifications that would make the basic one-epicycle one-deferent technique precisely fit the observed motion of the planets.

In antiquity the greatest of these attempts was made around A.D. 150 by the astronomer Ptolemy. Because his work displaced that of his predecessors and because all of his successors, including Copernicus, modelled their work upon his, the whole series of attempts for which Ptolemy provides the archetype is now usually known as Ptolemaic astronomy. That phrase, "Ptolemaic astronomy," refers to a traditional approach to the problem of the planets rather than to any one of the particular putative solutions suggested by Ptolemy himself, his predecessors, or his successors. Each of the particular individual solutions, and especially Ptolemy's, has an intense interest, both technical and historical, but both the particular solutions and their historical interrelationships are too complex to be considered here. Instead of attempting a general developmental account of the various Ptolemaic planetary systems, we shall therefore simply survey the main sorts of modification to which the basic epicycle-deferent system was subjected at various times between its first invention three centuries before Christ and its rejection by the followers of Copernicus.

Though their most important application is to the complex motions of the planets, the principal ancient and medieval modifications of the epicycle-deferent system are most simply described in their occasional applications to the apparently simpler motions of the sun and moon. The sun, for example, does not retrogress, so its motion does not require

a major epicycle of the sort described in the last section. But fixing the sun on a deferent that rotates uniformly about the earth as center does not give a quantitatively precise account of the solar motion, for, as shown by a reëxamination of the dates of the solstices and equinoxes listed in Chapter 1, the sun takes almost 6 days longer to move from the vernal equinox to the autumnal equinox (180° along the ecliptic) than it does to move back from the autumnal equinox to the vernal equinox (again 180°). The sun's motion along the ecliptic is slightly more rapid during the winter than summer, and such a motion cannot be produced by a fixed point on a uniformly rotating earth-centered circle. Examine Figure 22*a*, in which the earth is shown at the center of a uniformly rotating deferent circle and in which the positions of the vernal and autumnal equinoxes on the sphere of the stars are indicated by the dashes *VE* and *AE*. Uniform rotation of the deferent will carry the sun, S, from *VE* to *AE* in the same time that it takes to carry it back from *AE* to *VE*, and this corresponds only approximately with observation.

Suppose, however, that the sun is removed from the deferent and placed on a small epicycle that rotates once westward while the deferent rotates once eastward. Eight positions of the sun in such a system are shown in Figure 22*b*. It is clear that the summer half of the deferent's rotation does not carry the sun the entire distance from *VE* to *AE* and that the winter half of the rotation carries the sun farther than the distance from *AE* to *VE*. So the effect of the epicycle is to increase the time spent by the sun in the 180° between *VE* and *AE* and to decrease

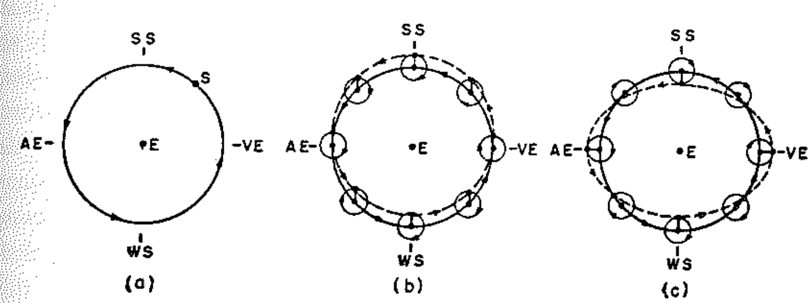

Figure 22. Functions of a minor epicycle. In (*a*) the sun, moved by a single earth-centered deferent, requires the same time to move from *AE* to *VE* that it needs to move back. In (*b*) the joint motion of deferent and minor epicycle carries the sun along the broken curve, so that more time is required for the trip from *VE* to *AE* than for the return. Diagram (*c*) shows the curve generated when the minor epicycle revolves at twice the rate used in constructing (*b*).

the time spent in the other half of the ecliptic between *AE* and *VE*. If the radius of the little epicycle is 0.03 the radius of the deferent, the difference in the time spent by the sun in the winter and the summer halves of the ecliptic will be the required 6 days.

The epicycle employed in the preceding discussion to correct a minor irregularity of the sun's motion is relatively small, and it produces no retrograde loops. Its function is therefore quite different from that of the larger epicycles considered in the last section, and, though Ptolemaic astronomers never did so, it will prove convenient to keep these two functions apart. Henceforth we shall use the term "major epicycle" for the large epicycles used to produce the qualitative appearance of retrograde motion and the term "minor epicycle" for the additional circles used to eliminate small quantitative discrepancies between theory and observation. All versions of the Ptolemaic system, both before and after Ptolemy, had just five major epicycles, and it is these with which Copernicus' reform did away. In contrast, the number of minor epicycles and similar devices needed to account for small quantitative discrepancies depended only on the precision of the available observations and on the accuracy of the predictions demanded from the system. The number of minor epicycles employed in the various versions of Ptolemaic astronomy therefore varied greatly from one version to the next. Systems employing half a dozen to a dozen minor epicycles were not uncommon in antiquity and the Renaissance, for by an appropriate choice of the size and speed of minor epicycles almost any sort of small irregularity could be explained away. That is why, as we shall see, Copernicus' astronomical system was so nearly as complex as Ptolemy's. Though his reform eliminated major epicycles, Copernicus was as dependent upon minor epicycles as his predecessors.

One sort of irregularity was treated with the aid of a minor epicycle in Figure 22*b*; another sort is shown in Figure 22*c*. There the minor epicycle rotates twice westward while the deferent moves once eastward. Combining the two rotations results in a total motion (broken line in the figure) along a flattened circle. A planet moving on this curve moves faster and spends less time in the vicinity of the summer and winter solstices than it does near the two equinoxes. If the epicycle had turned slightly less than twice while the deferent rotated once, then the positions on the ecliptic at which the planet's apparent speed was greatest would have changed on successive trips around the ecliptic. If it had appeared fastest near the summer solstice on one trip, it would have passed the summer solstice before gaining its greatest speed on the next trip. Other variations of this sort can be produced at will.

Uses of the minor epicycle are not limited to the nonretrogressing planets, the sun and moon. A minor epicycle can be placed upon a major epicycle and used in the prediction of the more elaborate planetary motions; in fact, planetary motions provided the minor epicycle's main astronomical application. One such application, an epicycle on an epicycle on a deferent, is shown in Figure 23*a*. If the major epicycle turns eight times eastward and the minor epicycle once westward during one rotation of the deferent, then the path within the sphere of the stars described by the planet is that shown in Figure 23*b*. It has eight normal retrograde loops, but these are somewhat more densely clustered in the half of the ecliptic between the vernal equinox and the autumnal equinox than in the half between the autumnal and vernal equinox. If the rate of rotation of the minor epicycle is now doubled, the path described by the planet is flattened as in Figure 22*c*. These diagrams begin to suggest the complexities of the paths that minor epicycles can produce.

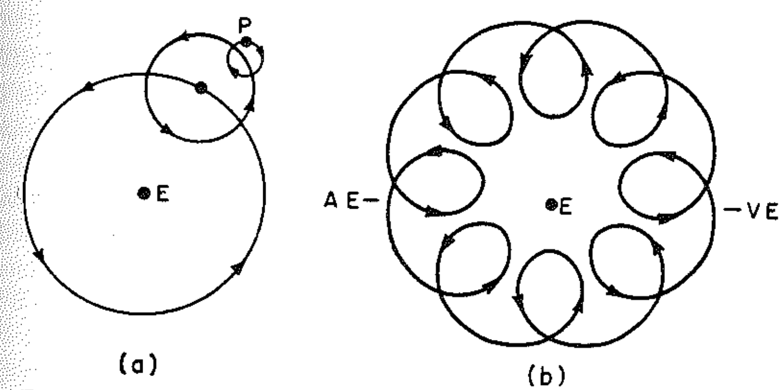

Figure 23. An epicycle on an epicycle on a deferent (*a*) and a typical path through space (*b*) generated by this system of compounded circles. For simplicity, the path has been shown rejoining itself smoothly, a situation that does not occur in the motion of real planets.

Nor is a minor epicycle the only device available for correcting minor discrepancies between one-epicycle one-deferent systems and the observed behavior of the planets. A glance at Figure 22*b* indicates that the effect there produced by a minor epicycle that rotates westward once as the deferent turns through a single eastward rotation can equally well be achieved by a single deferent whose center is displaced from the center of the earth. Such a displaced circle, known to ancient astronomers as an eccentric, is shown in Figure 24*a*. If the distance

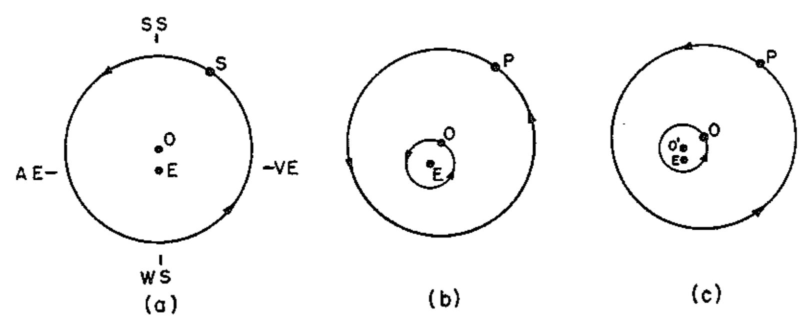

Figure 24. An eccentric (*a*), an eccentric on a deferent (*b*), and an eccentric on an eccentric (*c*).

between the earth, *E*, and the center of the eccentric, *O*, is about 0.03 the radius of the eccentric, this displaced circle will account for the 6 extra days that the sun spends between the vernal and autumnal equinoxes. That is the particular device that Ptolemy used in his own account of the sun. Other values of the distance *EO*, employed in conjunction with one or more epicycles, will account for other minor planetary irregularities. Additional effects may be obtained by placing the center of the eccentric on a small deferent (Figure 24*b*), or on a second smaller eccentric (Figure 24*c*). These two devices can be shown to be geometrically fully equivalent to a minor epicycle on a deferent and a minor epicycle on an eccentric, respectively, and most Ptolemaic astronomers used these small central circles in preference to minor epicycles. In all cases one or more epicycles may be added, and any or all of these circles may be tilted into different planes to account for the north and south deviations of the planets from the ecliptic.

One more device, the equant, was developed in antiquity to aid in the reconciliation of the theory of epicycles with the results of accurate observation. This device is of particular importance because Copernicus' aesthetic objections to it (Chapter 5) provided one essential

motive for his rejection of the Ptolemaic system and his search for a radically new method of computation. Copernicus used epicycles and eccentrics like those employed by his ancient predecessors, but he did not use equants, and he felt that their absence from his system was one of its greatest advantages and one of the most forceful arguments for its truth.

One form of equant, designed, for simplicity of illustration, to account for the previously discussed irregularity in the sun's motion, is shown in Figure 25. The center of the sun's deferent coincides as before with the center of the earth, *E*, but the deferent's rate of rotation is now required to be uniform not with respect to its geometric center *E*, but with respect to an equant point, *A*, displaced in this case toward the summer solstice. That is, the angle *a* subtended at the equant point *A* by the sun and the summer solstice is required to change at a constant rate. If the angle increases by 30° in one month, then it must increase by 30° in every month of the same length. In the figure the sun is shown at the vernal equinox, *VE*. To reach the autumnal equinox, *AE*, it must complete a semicircle, which will change the angle *a* by more than 180°, and to return from *AE* to *VE* it must complete a second semicircle, which will change *a* by less than 180°. Since every 180° increase in *a* requires the same amount of time, the sun must take longer to go from *VE* to *AE* than it requires for the return journey from *AE* to *VE*. Therefore, viewed from the equant point *A*, the sun travels at an irregular rate, fastest near the winter solstice and slowest near the summer solstice.

That is the defining feature of the equant. The rate of rotation of a deferent or some other planetary circle is required to be uniform, not with respect to its own geometric center, but with respect to an equant point displaced from that center. Observed from the geometric center of its deferent, the planet seems to move at an irregular rate or to wobble. Because of the wobble, Copernicus felt that the equant was not a legitimate device for application to astronomy. For him the apparent irregularities of the rotation were violations of the uniform circular symmetry that made the system of epicycles, deferents, and eccentrics so plausible and attractive. Since the equant was normally applied to eccentrics and since similar devices occasionally made the epicycle wobble as well, it is not hard to imagine how Copernicus might have considered this aspect of Ptolemaic astronomy monstrous.

The mathematical devices sketched in the preceding pages were not all developed at a stroke or by Ptolemy. Apollonius, in the third century B.C., knew both major epicycles (Figure 19*a*) and eccentrics with moving centers (Figure 24*b*). During the following century Hipparchus added minor epicycles and a more general theory of eccentrics to the arsenal of astronomical weapons. In addition he combined these devices to provide the first quantitatively adequate account of the irregularities in the motions of the sun and moon. Ptolemy himself added the equant, and during the thirteen centuries between his time and that of Copernicus, first Moslem and then European astronomers employed still other combinations of circles – including the epicycle on an epicycle (Figure 23*a*) and the eccentric on an eccentric (Figure 24*c*) – to account for additional planetary irregularities.

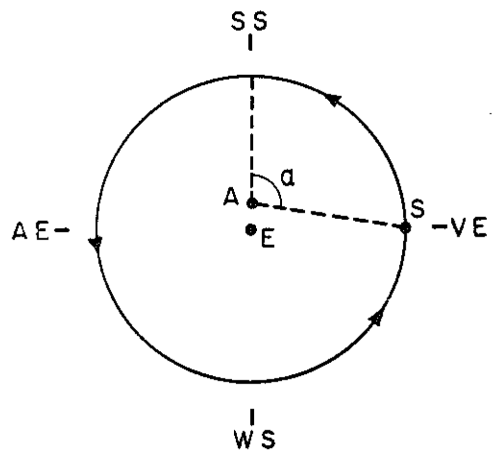

Figure 25. The equant. The sun, S, moves on the earth-centered circle but at an irregular rate determined by the condition that the angle α vary uniformly with time.

But Ptolemy's contribution is the outstanding one, and this entire technique of resolving the problem of the planets is appropriately known by his name, because it was Ptolemy who first put together a particular set of compounded circles to account, not merely for the motions of the sun and moon, but for the observed quantitative regularities and irregularities in the apparent motions of all the seven planets. His *Almagest*, the book that epitomizes the greatest achievements of ancient astronomy, was the first systematic mathematical treatise to give a *complete, detailed,* and *quantitative* account of all the celestial motions. Its results were so good and its methods so powerful that after Ptolemy's death the problem of the planets took a new form. To increase the accuracy or simplicity of planetary theory Ptolemy's successors added epicycles to epicycles and eccentrics to eccentrics, exploiting all the immense versatility of the fundamental Ptolemaic technique. But they seldom or never sought fundamental modifications of that technique. The problem of the planets had become simply a problem of design, a problem to be attacked principally by the rearrangement of existing elements. What particular combination of deferents, eccentrics, equants, and epicycles would account for the planetary motions with the greatest simplicity and precision?

We cannot pursue further the individual quantitative solutions of this problem proposed by Hipparchus, Ptolemy, and their successors. The complete quantitative systems are mathematically too complex. Much of Ptolemy's *Almagest* consists of quantitative mathematical tables, diagrams, formulas, and proofs, of long illustrative computations, and of lists of numerous observations. Yet the problems that set Copernicus to searching for a new approach to the problem of the planets and the advantages that he claimed to derive from his new system all lie within this abstruse body of quantitative theory. Copernicus did not attack the two-sphere universe, though his work ultimately overthrew it, and he did not abandon the use of epicycles and eccentrics, though these too were abandoned by his successors. What Copernicus did attack and what started the revolution in astronomy was certain of the apparently trivial mathematical details, like equants, embodied in the complex mathematical systems of Ptolemy and his successors. The initial battle between Copernicus and the astronomers of antiquity was fought over technical minutiae like those sketched in this section.

The Anatomy of Scientific Belief

For its subtlety, flexibility, complexity, and power the epicycle-deferent technique sketched in the two preceding sections has no parallel in the history of science until quite recent times. In its most developed form the system of compounded circles was an astounding achievement. *But it never* quite *worked.* Apollonius' initial conception solved the primary planetary irregularities – retrograde motion, variation of brightness, alteration in the time required for successive journeys around the ecliptic – and it did so simply and at a stroke. But it also disclosed some residual secondary irregularities. Some of these were explained away by the more elaborate system of compounded circles developed by Hipparchus, but still the theory did not quite match the results of observation. Even Ptolemy's complex combination of deferents, eccentrics, epicycles, and equants did not precisely reconcile theory and observation, and Ptolemy's was neither the most complex nor the last version of the system. Ptolemy's many successors, first in the Moslem world and then in medieval Europe, took up the problem where he had left it and sought in vain for the solution that had evaded him. Copernicus was still grappling with the same problem.

There are many variations of the Ptolemaic system besides the one that Ptolemy himself embodied in the *Almagest*, and some of them achieved considerable accuracy in predicting planetary positions. But the accuracy was invariably achieved at the price of complexity – the addition of new minor epicycles or equivalent devices – and increased complexity gave only a better approximation to planetary motion, not finality. No version of the system ever quite withstood the test of additional refined observations, and this failure, combined with the total disappearance of the conceptual economy that had made cruder versions of the two-sphere universe so convincing, ultimately led to the Copernican Revolution.

But the Revolution was an incredibly long time coming. For almost 1800 years, from the time of Apollonius and Hipparchus until the birth of Copernicus, the conception of compounded circular orbits within an earth-centered universe dominated every technically developed attack upon the problem of the planets, and there were a great many such attacks before Copernicus'. Despite its slight but recognized inaccuracy and its striking lack of economy (contrast the earlier two-sphere universe described in Chapter 1), the developed Ptolemaic system had an immense life span, and the longevity of this magnificent but clearly imperfect system poses a pair of closely related puzzles: How did the two-sphere universe and the associated epicycle-deferent planetary theory gain so tight a grip upon the imagination of the astronomers? And, once gained, how was the psychological grip of this traditional approach to a traditional problem released? Or to put the same ques-

tions more directly: Why was the Copernican Revolution so delayed? And how did it come to pass at all?

These are questions about the history of a particular set of ideas, and as history they will be considered at some length below. But they are also more generally concerned with the nature and structure of conceptual schemes and with the process by which one conceptual scheme replaces another. It is therefore illuminating to approach them first by returning briefly to the abstract logical and psychological categories introduced in the penultimate section of the first chapter. We there examined the functions of a conceptual scheme: we now ask how a smoothly functioning scheme, like the early two-sphere universe, can be replaced. Examine the logic of the phenomenon first.

Logically there are always many alternative conceptual schemes capable of bringing order to any *prescribed* list of observations, but these alternatives differ in their predictions about phenomena not included on the list. Both the Copernican and the Newtonian systems will account for naked-eye stellar and solar observations just as adequately as will the two-sphere system; Heraclides' system will do the same and so will the system developed by Copernicus' successor, Tycho Brahe; in theory there are an infinite number of other alternatives besides. But these alternatives agree principally about observations that have already been made. They do not give identical accounts of all possible observations. The Copernican system, for example, differs from the two-sphere universe in predicting an apparent annual motion of the stars, in demanding a much larger diameter for the stellar sphere, and in suggesting (though not to Copernicus) a new sort of solution for the problem of the planets. It is because of differences like these (and there are many others besides) that a scientist must believe in his system before he will trust it as a guide to fruitful investigations of the *unknown*. Only one of the different alternatives can *conceivably* represent reality, and the scientist exploring new territory must feel confident that he has chosen that one or the closest of the available approximations to it. But the scientist pays a price for this commitment to a particular alternative: he may make mistakes. A single observation incompatible with his theory demonstrates that he has been employing the wrong theory all along. His conceptual scheme must then be abandoned and replaced.

That, in outline, is the logical structure of a scientific revolution. A conceptual scheme, believed because it is economical, fruitful, and cosmologically satisfying, finally leads to results that are incompatible with observation; belief must then be surrendered and a new theory adopted; after this the process starts again. It is a useful outline, because the incompatibility of theory and observation is the ultimate source of every revolution in the sciences. But historically the process of revolution is never, and could not possibly be, so simple as the logical outline indicates. As we have already begun to discover, observation is never *absolutely* incompatible with a conceptual scheme.

To Copernicus the behavior of the planets was incompatible with the two-sphere universe; he felt that in adding more and more circles his predecessors had simply been patching and stretching the Ptolemaic system to force its conformity with observations; and he believed that the very necessity for such patching and stretching was clear evidence that a radically new approach was imperatively required. But Copernicus' predecessors, to whom exactly the same sorts of instruments and observations were available, had evaluated the same situation quite differently. What to Copernicus was stretching and patching was to them a natural process of adaptation and extension, much like the process which at an earlier date had been employed to incorporate the motion of the sun into a two-sphere universe designed initially for the earth and stars. Copernicus' predecessors had little doubt that the system would ultimately be made to work.

In short, though scientists undoubtedly do abandon a conceptual scheme when it seems in irreconcilable conflict with observation, the emphasis on logical incompatibility disguises an essential problem. What is it that transforms an apparently temporary discrepancy into an inescapable conflict? How can a conceptual scheme that one generation admiringly describes as subtle, flexible, and complex become for a later generation merely obscure, ambiguous, and cumbersome? Why do scientists hold to theories despite discrepancies, and, having held to them, why do they give them up? These are problems in the anatomy of scientific belief. They are the primary concern of the next two chapters, which set the stage for the Copernican Revolution proper.

Our immediate problem, however, is the analysis of the grip exerted upon men's minds by the ancient tradition of astronomical research. How could this tradition provide a set of mental grooves that guided the astronomical imagination, limited the conceptions available in research, and made certain sorts of innovations difficult to conceive and more difficult to accept? We have already dealt, at least implicitly, with the strictly astronomical aspects of this problem. Both the two-sphere universe and the associated epicycle-deferent technique were initially highly economical and fruitful; their first successes seemed to guarantee the fundamental soundness of the approach; surely only minor modifications would be required to make the mathematical predictions correspond with observation. A conviction of this sort is difficult to break, particularly once it has been embodied in the practice of a whole generation of astronomers who transmit it to their successors through their teaching and writing. This is the band-wagon effect in the realm of scientific ideas.

The band-wagon effect is not, however, the whole explanation of the strength of the astronomical tradition, and in trying to complete the explanation we shall be temporarily led away from astronomical problems altogether. The two-sphere universe provided a fruitful guide to the solution of problems outside as well as inside astronomy. By the end of the fourth century B.C. it had been applied not only to the problem of the planets, but also to terrestrial problems, like the fall of a leaf and the flight of an arrow, and to spiritual problems, like the relation of man to his gods. If the two-sphere universe, and particularly the conception of a central and stable earth, then seemed the indubitable starting point of all astronomical research, this was primarily because the astronomer could no longer upset the two-sphere universe without overturning physics and religion as well. Fundamental astronomical concepts had become strands in a far larger fabric of thought, and the nonastronomical strands could be as important as the astronomical in binding the imagination of astronomers. The story of the Copernican Revolution is not, therefore, simply a story of astronomers and the skies.