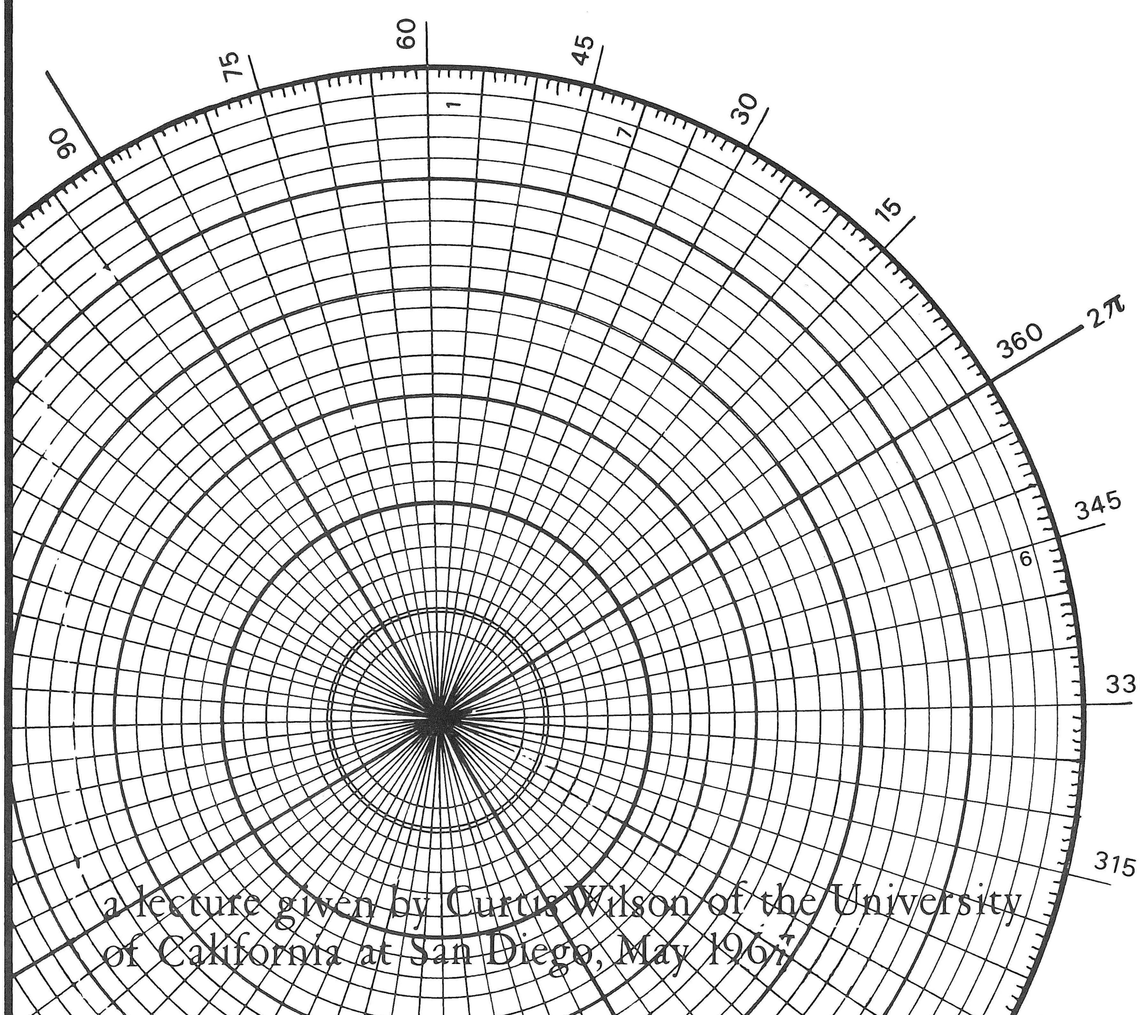
KEPLER'S DISCOVERY OF THE ELLIPTICAL ORBIT OF MARS



a lecture given by Curtis Wilson of the University of California at San Diego, May 1967

For a good many years I have had the ambition to be able to answer the question: How did Kepler arrive at the so-called first Keplerian law, the ellipticity of the planetary orbit? The discovery of the ellipse ought to be, one would think, a rather straightforward geometrical affair, a matter of calculating distances between Sun and planet trigonometrically, and seeing that these distances fit into an ellipse with the Sun at one focus. Now as all the commentators make clear, such distance-determinations do indeed play a role in Kepler's journey of discovery. But this role, as I believe I can now show, is a negative one: the trigonometry does not lead to the idea of the ellipse in the first place, it is too imprecise to give the dimensions of the ellipse, and it does not provide what Kepler regarded as an adequate confirmation of the elliptical shape. What has emerged for me from a study of Kepler's book, the *Astronomia Nova* of 1609 -- and I do not believe that this understanding appears in any of the secondary literature -- is not merely that Kepler goes his journey theory-laden, but that it is only thus that he manages to arrive at his first two laws, the ellipse and the area law. It is an initial hunch, a physical hypothesis, which guides him throughout; every step is made deliberately, not only in confrontation with data, but in pursuance of that hunch. And the laws arrived at, the ellipse and the area law, rest on the physical theory as on a premise; they remain hypothetical; they cannot be confirmed independently of one another, but taken jointly yield better predictions than had ever been achieved before. What I shall try to show, then, is how this is so; to trace in schematic outline the logical map of Kepler's journey.

Let me begin with two preliminary remarks. The first has to do with the observations that Kepler used. With the benefit of hindsight, we can say that all the planets known to Kepler have nearly circular orbits. The one with the most flattened orbit is Mercury, but Mercury, being close to the Sun, is difficult to observe. The next greatest departure from circularity occurs in the orbit of Mars -- the orbit that Kepler studies -- and here the departure from circularity is such that the minor axis of the ellipse is about 1/200 or ½% less than the major axis. In a picture of this orbit drawn to scale, most of us could not, by merely looking, detect the departure from circularity.

Thus a planetary theory which uses circular paths can be surprisingly good. It is not quite right to talk about errors arising from the assumption of circularity alone; in any planetary theory, an additional assumption has to be made as to the motion of the planet. But I can assert that, if one did have a theory for Mars which erred only in assuming a circular path, the discrepancies between prediction and observation would never rise above 10' of arc. Now my thumb, held out as far as possible from my eye, subtends an angle of about two degrees, or 120'. So the discovery of the elliptical orbit is going to depend on rather refined observations.

The observations which Kepler used were made by Tycho Brahe, in the years between 1575 and 1600. They were made by naked eye, but with large instruments, quadrants, sextants, on which the scales could be finely divided. Often Tycho had two observers observing the same object simultaneously; their results sometimes differed by as much as 3' of arc, but the final result or average was regarded as accurate to within 2' of arc. This is to be compared with the 10' discrepancy between observations which previous astronomers had allowed to be tolerable. Without Tycho's reduction of the expected error to about 2' of arc, Kepler's discoveries could not have been made.

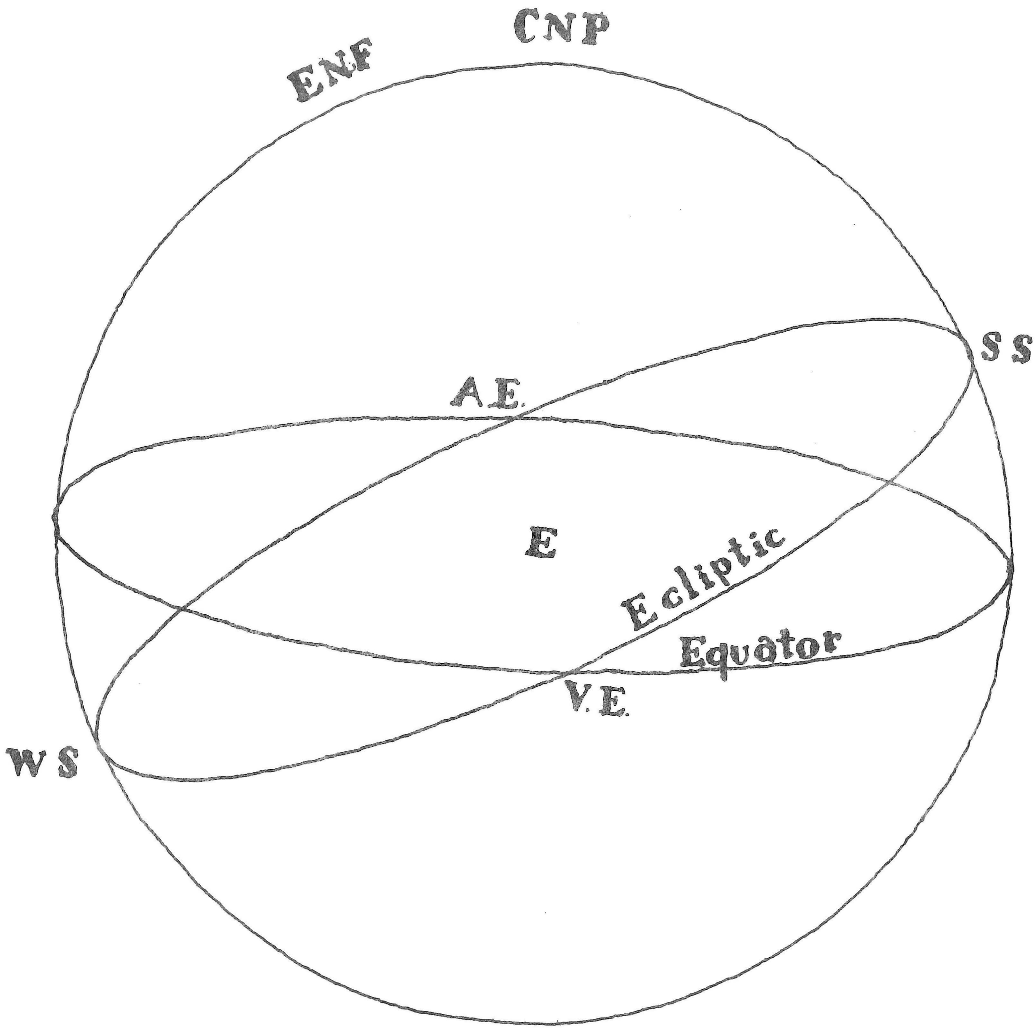


FIGURE 1 STAGE 1

The second preliminary remark concerns Kepler's hunch, the physical hypothesis which guides him throughout his work. When Kepler went to Prague in 1600, to work under Tycho Brahe, he was already an ardent Copernican. He saw that the preeminent role of the Sun in planetary theory, which had appeared as an unexplained assumption in Ptolemy's constructions, becomes a simple consequence of the postulates in Copernicus's system. However, Kepler felt that Copernicus had failed to realize the full meaning of his new system. For Copernicus, the Sun sits like a lamp in the midst of the solar system, lighting and heating up the world; otherwise it is functionless. The centers of the planetary circles do not even lie in the sun.

Now Kepler felt that the Sun was more then a mere lamp. He noted that as you go out from the Sun, the actual linear speeds of the planets become less: Venus moves more slowly than Mercury, the Earth moves more slowly than Venus, Mars moves more slowly than the Earth, and so on. To Kepler this correlation between distance and speed suggested a causal relation: he hypothesized that the Sun was somehow causing the planets to move about, and that the motive force or virtus fell off in strength as one got farther from the Sun. Thus, when he started his work on Mars, Kepler already had the germ of the idea of a celestial physics, that is, an account of the planetary movements in which the planets do not have within them the source of their motions, but are moved from without by pushes and pulls. What I think we need to see is how Kepler, faced with a mass of confusing data and theory inherited from his predecessors, is guided by this physical hypothesis throughout his journey.

So much for preliminaries. I turn now to what Kepler, with his flair for the dramatic, calls his war on Mars. I shall divide the war into seven phases.

The first phase I shall barely mention. It has to do with the latitudes of the planet. The observed position of a planet is specified by latitude and longitude (Figure 1). Here is the celestial sphere, with the celestial north pole on top, and the equator running around the middle. As far as the making of observations is concerned, we can think of the Earth as being at or near the center of this sphere. Inclined to the equator is the circle of the ecliptic. This is the apparent path of the Sun during a year, projected against the background of the stars. The Sun's position along the ecliptic, measured in degrees eastward from the vernal equinox, is called its longitude.

Now the planets are observed to do something like what the Sun does: their general movement is eastward, approximately along the ecliptic. However, they depart from the ecliptic by small amounts, going above it and below. Now the theories of Ptolemy and Copernicus with regard to the variation in latitudes are frightfully complicated. Kepler accomplishes an enormous simplification. He hypothesizes that the plane of Mar's orbit is inclined at a constant angle to the plane of the ecliptic, and passes through the Sun. He is able to verify this hypothesis by means of certain observations, which I shall not describe here. Note the use of the Sun as a fixed reference point; the previous theories had always used an imaginary point called the Mean Sun, rather than the "real and apparent Sun", as a reference point in their planetary calculations, and this was one of the reasons for the complications in the Ptolemaic and Copernican theories of the latitudes. (In the case of Ptolemy, the additional complication due to the use of the Mean Sun is clearly apparent in the theories of Venus and Mercury; in the Ptolemaic theories of the latitudes of the superior planets, the complication arises mainly from the geocentric mode of description.) Kepler's establishment

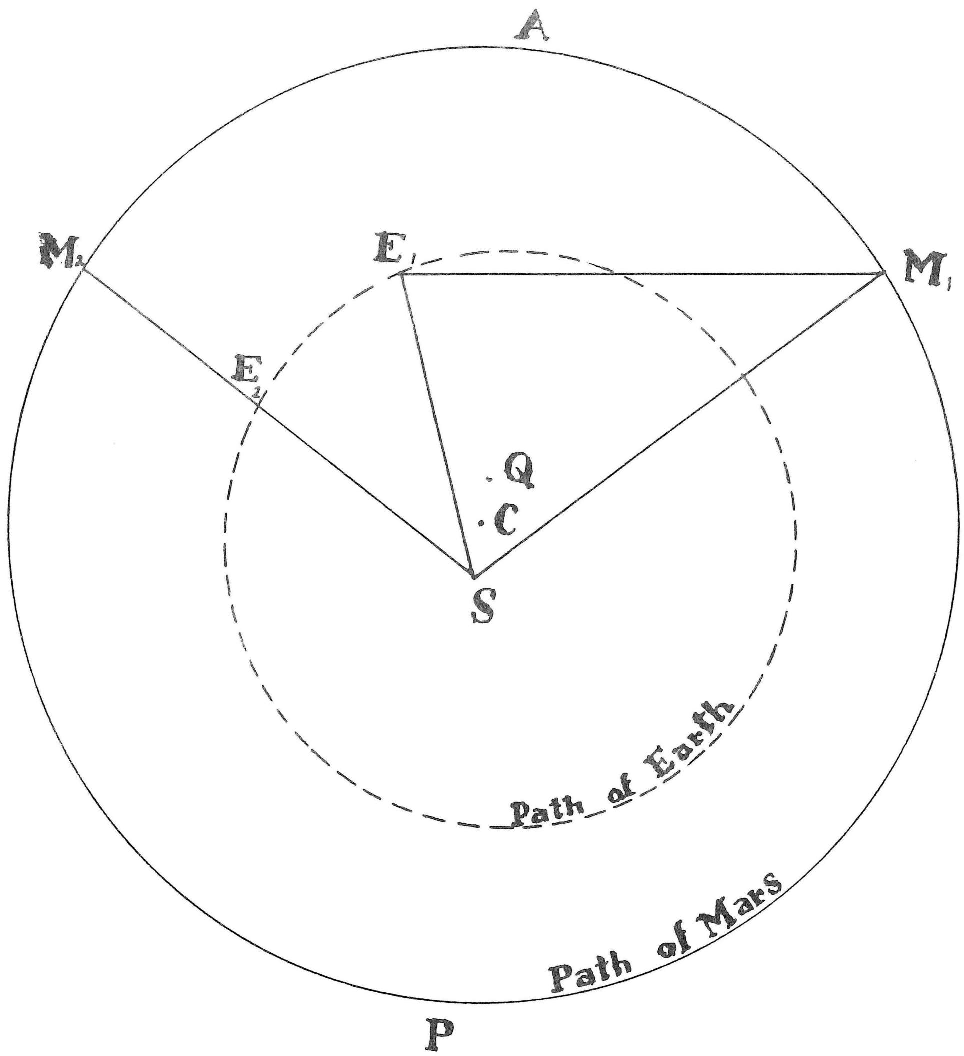


FIGURE 2 STAGE 2

M_1 = Mars as Evening Star
M_2 = Mars at Opposition

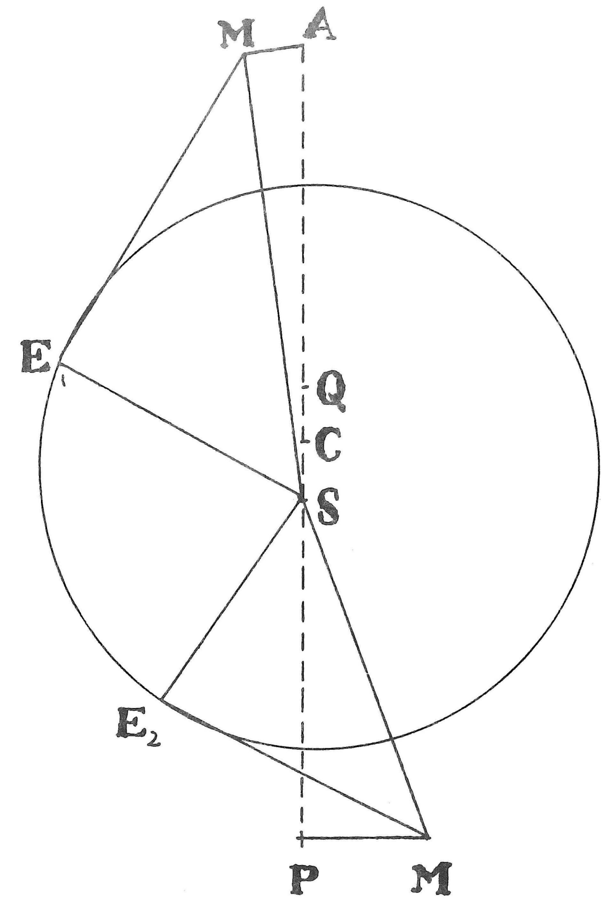
of the constancy of the inclination, and the passage of the plane of the orbit through the real Sun, is a first victory for his hunch about the role of the Sun. In what follows, I shall ignore the problem of latitudes entirely. I shall deal only with the longitudes, i.e., the positions of Mars as projected onto the ecliptic.

Second Phase. Kepler sets up a theory for the longitudes of Mars which is successful in one way and fails in another. The first thing to understand is what one has to go on in setting up a theory for the longitudes of a planet. The observations of the planet are of two kinds. Let us agree to take the Copernican standpoint, and to think of the Sun as a fixed point, and of the Earth and the other planets as moving in orbits about the Sun (Figure 2). Here the outer circle is the assumed path of Mars, and the inner dotted circle is the assumed path of the Earth. Ordinarily, when an observation on Mars,is made, the Earth, Sun and Mars form a triangle, e.g., SE_1M_1. The observation gives only the direction of Mars from the Earth, against the background of the stars. At this point we are far from knowing everything about the triangle. Of course Kepler does have a theory of the Earth's motion, or rather of the Sun's motion, which he has inherited from Tycho, and we shall find him trusting this theory to give the direction of the line E_1S_1. For a reason which will appear shortly, Kepler will not completely trust this theory to give him the distance E_1S_1, which varies around the circle since the Sun is off-center. And even if we accepted Tycho's theory, to solve the triangle we would still need to know the direction of SM_1. How can we determine the direction of SM_1?

This brings us to the second sort of observation. About every 780 days, Mars, the Earth, and the Sun come into a line, with the Sun on the far side of the Earth from Mars. At about this time, Mars can be observed approximately on the meridian overhead at midnight. At the exact time of opposition, an observer on Earth is seeing Mars against the background of the stars in just the position in which he would see it if he were situated in the Sun. Kepler had twelve such observations to work with, taken between 1580 and 1604. Actually, one seldom gets an observation at the exact time of the opposition; the position of the planet at opposition has to be calculated from a group of observations made about the time of the opposition. Kepler, for the first time, calculates the oppositions to the real Sun, rather than the Mean Sun; this is in line with his hunch; he hopes this change will lead him to a new and better theory.

But how set up a theory? At the time of the opposition one knows the heliocentric longitude of Mars, its position on the ecliptic as seen from the Sun. But one does not know its distance from the Sun. The only thing to do is to make an assumption. Kepler proceeds to try to fit a theory of Ptolemaic type to the data. That is, he assumes that Mars is moving in a circular path with center at C, that the Sun is off-center at some point S, and that the motion of the planet is uniform not necessarily about C but about another point Q, called the equant point. All of these features are present in the Ptolemaic theory of the deferent of Mars, except of course that the Sun is replaced by the Earth. What one needs to find is the direction of the line SCQ, and the ratios of SC and CQ to the radius of the circle. The procedure for finding these things, both for Ptolemy and Kepler, is the horrendous one of trial and error: make a guess, then alter it if the theory fails to fit the facts.

There is one difference between Ptolemy and Kepler that I should mention: Ptolemy had assumed that the point C is midway between S and Q for reasons that



By vicarious theory:

SQ = 18,564
½SQ = 9,282
SC = 11,332

From two trigonometric calculations:

SC between
8000 and 9943
(av.= 8971)
SC between
8377 and 10106
(av.= 9242)

FIGURE 3 STAGE 2

are not entirely clear. Kepler did not want to assume this; rather, he wanted to place the point C between S and Q at just the right point, to yield the best fit between observation and theory.

Kepler uses four of the twelve observations at opposition to set up his theory, and after 70 trials, finds a theory which fits these observations. He then checks the theory against the remaining eight observations, and finds the concordance very good: the average discrepancy between theory and observation is 50", the maximum discrepancy being 2'12". What this theory does then, is to give the heliocentric longitudes of Mars, its positions as seen from the Sun, with a precision equal to that of the observations.

Now as we next discover, this theory is false. Kepler comes to call it his _vicarious_ or substitute theory, because though false it serves to give him the heliocentric longitudes of Mars, for all his later work. What it fails to give correctly are the _distances_ of Mars from the Sun.

How can Kepler learn _anything_ about the _distances_ of Mars from the Sun? Suppose for a moment that we take Tycho's solar theory and transform it into a theory of the Earth's motion -- this is a very simple geometrical transformation which I shall describe shortly -- and consider observations of Mars when it is _not_ in opposition (Figure 3). The Earth, Sun and Mars in each of the two observations pictured here form a triangle. The position of SM is now known from the vicarious theory, the position of EM is given by the observation in each case, and the position of ES is determined by Tycho's solar theory. Now Tycho's solar theory, which differs from Ptolemy's only in its numbers, is known to give the heliocentric longitudes of the Earth with considerable accuracy, just as the vicarious theory gives the heliocentric longitudes of Mars with considerable accuracy, so all the angles in these triangles are known to within about 4' of arc. The ratios of the sides of the triangles can then be calculated trigonometrically. If only one could trust Tycho's theory to give also the ratio of SE_1 to SE_2 -- these are not equal because the Sun is not in the center of the circle -- one could determine the ratio of SM_1 to SM_2.

The fact is, Kepler knows that he can trust Tycho's theory, in its prediction of Earth-Sun distances, within certain limits. One basis for this trust is provided by observations of the apparent diameter of the Sun's disk, which show, though a little roughly, that the distance between the Earth and Sun does not change very much, changes in fact less than Tycho's theory predicts. The orbit _must_ be nearly circular. The possible errors are small enough so that Kepler can use Tycho's theory to learn something about Mars-Sun distances. He is particularly interested in checking the placement of the point C, the center of the orbit. The vicarious theory had put it about 0.61 of the way from S to Q. By computing Mars-Sun distances near the line of apsides Kepler now shows that C must be much closer to half way between S and Q. If he now alters the vicarious theory accordingly, and puts C midway between S and Q, the altered theory no longer predicts the heliocentric longitudes of Mars correctly, but gives errors as high as 8', whereas Tycho's observations cannot be in error by more than about 2'. The 8' error here, Kepler tells us, is what forced him to go on to a total reformation of astronomy.

Kepler's situation, at this point, can be summarized as follows. He has two theories, Tycho's solar theory for the Earth, and the vicarious theory for

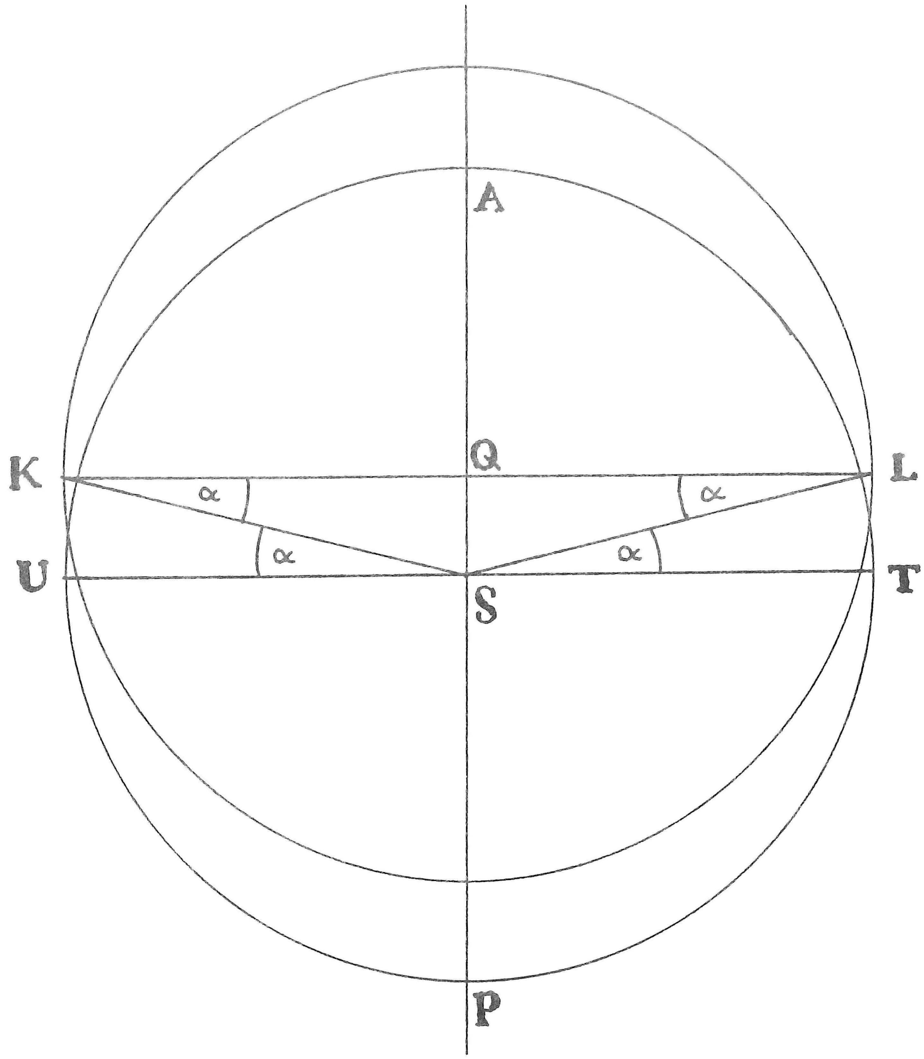


FIGURE 4 STAGE 2

Why the Vicarious Theory succeeds for heliocentric longitudes

Mars. Both are Ptolemaic-style theories, involving a circular path, and uniform angular motion about a point within the circle. Each theory predicts the heliocentric longitudes of its planet with errors not exceeding 1' or 2', as confirmed rather directly by observation.* But if Tycho's theory for the Earth is taken as correct for Earth-Sun distances, then the vicarious theory is wrong with respect to Mars-Sun distances. In fact, the vicarious theory is definitely wrong, because the possible errors in the distances in Tycho's theory are relatively small, compared with those deduced for the vicarious theory. So of the two assumptions on which the vicarious theory rests, the circular path and the uniform motion about an equant point, one or the other must be wrong. The vicarious theory will remain useful, and indeed indispensable, for finding the heliocentric longitudes of Mars, but it is a false theory.

Which of the two Ptolemaic principles, circular path or equant point, should be changed? Kepler has no doubts on this score. If one has in mind the possibility of a celestial physics, the equant principle appears artificial: there is no body at the equant point, no believable mechanism of which the equant principle would be the expression. Moreover, the work on the vicarious theory has provided Kepler with a clue to a possible substitute for the equant principle. It is the equant principle that Kepler will first abandon, but before doing so, he undertakes an inquiry as to why Ptolemaic-style theories, using equant points, can be so successful, in the prediction of heliocentric longitudes.

Fundamentally, the success means this (Figure 4). The varying motion of any planet is symmetrical about a certain line, the line of apsides (AP). The most rapid motion of the planet occurs at one point (P), and its slowest motion at a point 180° away (A). In between these points, on either side of the line of apsides, the change in the planet's angular speed about S is gradual. Suppose we assume the planet to be moving uniformly on a circle concentric with S. This theory is wrong, because, starting at A, the point of slowest motion, the actual planet falls behind the planet of our theory; the discrepancy builds up gradually to a maximum at about a quarter of the planet's period, and then decreases again to zero as the planet comes to P.

Suppose that after a quarter period the planet is seen from S along SK instead of SU; in other words, the discrepancy at this time between observation and our first theory is the angle KSU. We can eliminate this discrepancy rather simply, by shifting the planetary circle up, so that its center is at Q. This new theory will rhyme with observation when the planet is in the lines SA, SK, SP, SL; and the large-order discrepancy between theory and observation, angle KSU (= angle SKQ), amounting to nearly 11° in the case of Mars, has been eliminated. Such discrepancies as remain will be in the octants, and will be much smaller; in the case of Mars these remaining errors are at maximum about 8', i.e., 1/80 of the former error.

Kepler goes on to show how, keeping Q fixed as the equant point or center of uniform motion, and taking a new orbit with center at some point C between S and Q, the remaining discrepancies can be reduced below the level of observational detection. In this final adjustment of the orbit, the point Q, previously deter-

*This statement is made from Kepler's point of view. Actually, Tycho's incorrect values for refraction and for solar parallax lead to an eccentricity that is too high, and so the solar theory is more erroneous than the above statement implies.

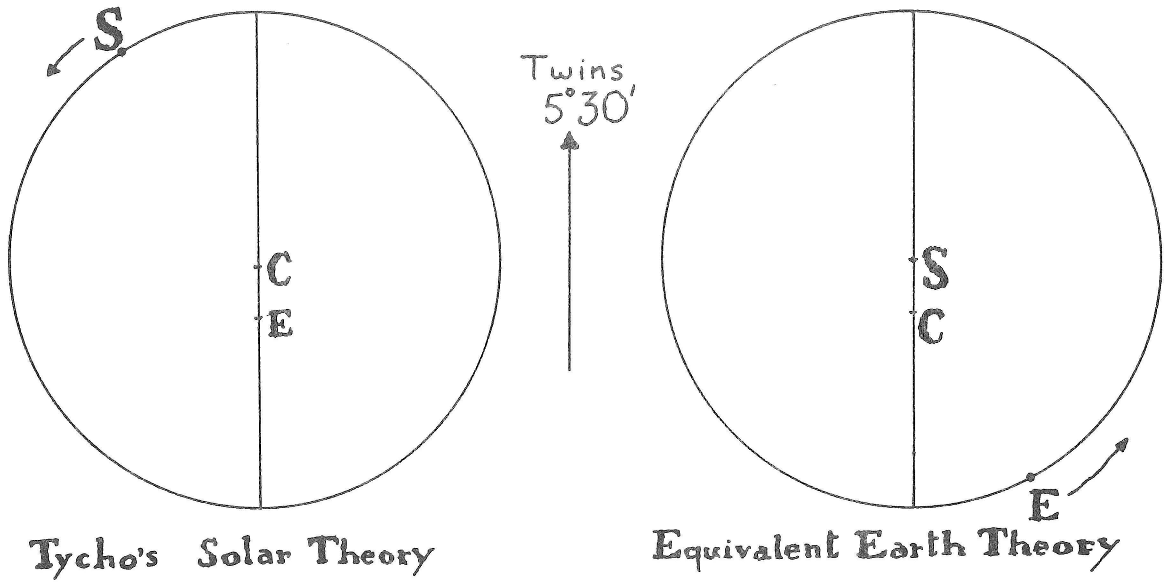
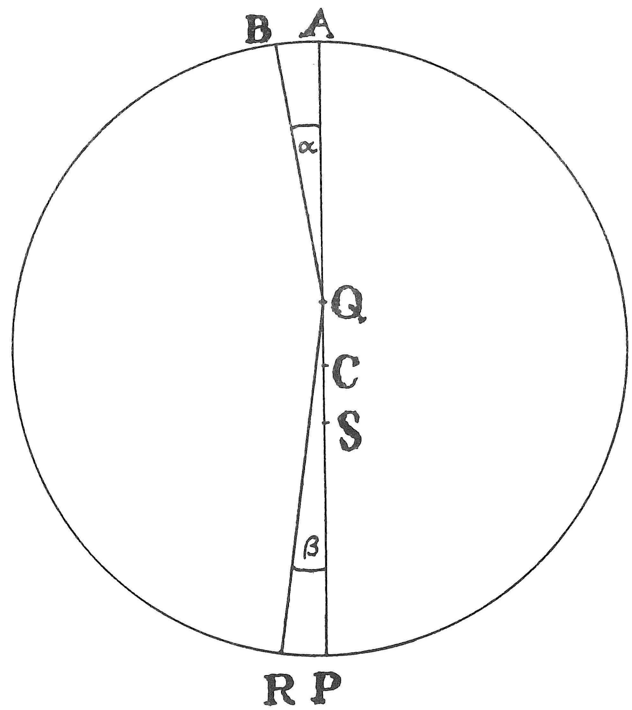


FIGURE 5 STAGE 3



Given:
(1) SC = CQ
(2) $\alpha = \beta$

Then:
AB:RP :: AQ:QP
(approximately)

But:
AQ = SP,
QP = SA.

Therefore:
AB:RP :: SP:SA
(approximately)

FIGURE 6 STAGE 3

mined, must remain the equant point; otherwise, some part of the original large-order discrepancy will be re-introduced.

This argument shows why, given the general character of planetary motion, a theory of the equant type is doomed to succeed, as far as concerns prediction of heliocentric longitudes. There is one other consequence I should like to mention for later use: it is that the placement of the equant point can be determined with greater precision than the placement of the center of the orbit, because the placement of the equant point depends on the determination of a larger angle.

<u>Third Stage</u>. Kepler devises a replacement for the equant principle. As I mentioned before, the work with Mars has already provided a clue. Kepler has found that the midpoint of the line of apsides, C, cannot be where the vicarious theory puts it, but must be more nearly midway between S and Q, the Sun and the equant point. Now it is a peculiar fact, on which Kepler had meditated, that Ptolemy in his theories of Venus, Mars, Jupiter, and Saturn, had assumed what is called the bisection of the eccentricity. If we translate the Ptolemaic theories into heliocentric form, what this means is that the center of the orbit is assumed to lie midway between the Sun and the Equant point. And Kepler has now gotten a rough verification of this assumption in the case of Mars. Among the principal planets, the main exception, to which this assumption had not been applied, was the Sun, or if you are a Copernican, the Earth (Figure 5). (Of tee other exception, Mercury, Kepler ends by distrusting some of the observations on which Ptolemy's theory is based.) Tycho's solar theory, like Ptolemy's, was a simple eccentric theory: the Earth is off-center, but the Sun goes uniformly about the center of the circle, so that equant point and center of the circle coincide. When you transform this theory into a theory of the Earth's motion, you again get a simple eccentric theory, with equant point and center of the orbit coinciding.

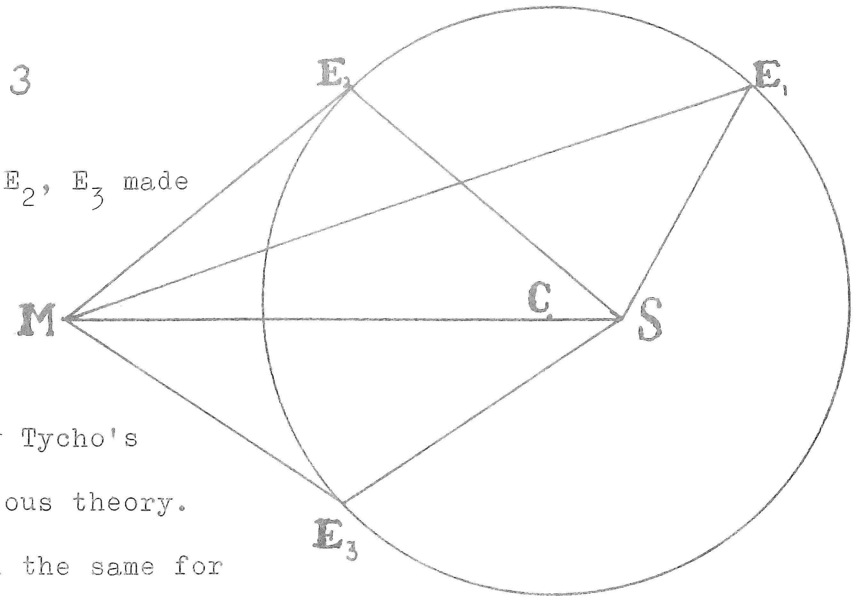
Now what if the eccentricity were bisected in the case of all the planets? Take a planetary theory in which the eccentricity is bisected (Figure 6). The planet moves at a uniform angular rate about Q. In a given time it would, say, go through a small angle AQB. In an equal time, later, it would go through the equal angel RQP. The equality of the angles means that, for small arcs near the line of apsides, very nearly, $\widehat{AB}/\widehat{RP} = AQ/QP$. But because of the bisection of the eccentricity, because Q and S are symmetrically placed about C, it follows that $AQ = SP$ and $AS = QP$. Substituting, $\widehat{AB}/\widehat{RP} = SP/SA$. In other words, near the line of apsides, the arcs traversed by the planet in equal times are inversely as the distances from the Sun. Or the velocity of the planet varies inversely as its distance from the Sun.

This is a new hypothesis, which Kepler believes might be generally applicable, throughout the orbit. It is not precisely equivalent to the equant principle, except in the line of apsides when the eccentricity is bisected. But this hypothesis is in accord with Kepler's hunch, the physical hypothesis he brought with him to Prague, the idea that the Sun, by means of some mysterious, immaterial virtue, is pushing the planets about, and that its motive virtue falls off in strength with greater distances from the Sun.

Now Kepler feels that he will be justified in trying out this new hypothesis, if he can establish that the eccentricity is bisected in the case of the Earth, just as Ptolemy had assumed it to be for the other planets besides Mercury.

FIGURE 7

STAGE 3



Observations at E_1, E_2, E_3 made 687 days apart.

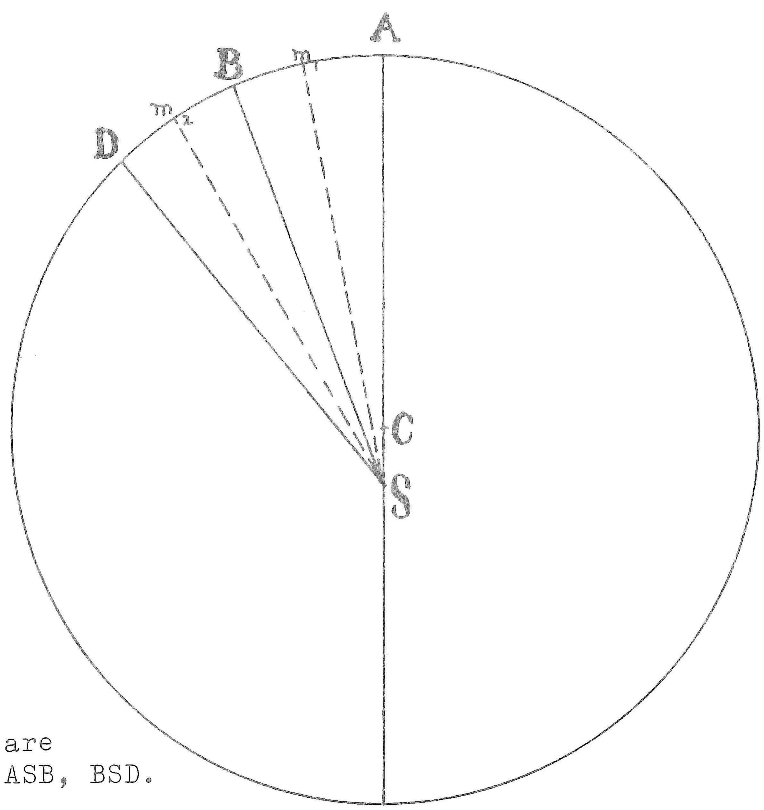
Directions of lines determined:

(1) E_1M, E_2M, E_3M by observation.
(2) SE_1, SE_2, SE_3 by Tycho's solar theory.
(3) SM by the vicarious theory.

Length of SM assumed the same for each observation.

FIGURE 8

STAGE 3



"Radius Rule":
Times for equal arcs AB, BD, are as the distances of those arcs from the Sen, Sm_1, Sm_2.

"Area Rule":
Times for equal arcs AB, BD are as the areas of the sectors ASB, BSD.

The main procedure Kepler uses is as follows (Figure 7). He takes three of Tycho's observations of Mars, made 687 days apart, the time it takes Mars to go once around its orbit. At all three times, then, Mars should be at the same place, M. Let the Earth at these times be at the points E_1, E_2, E_3. Then the directions of the lines E_1M, E_2M, E_3M are determined by Tycho's observations; the directions of SE_1, SE_2, SE_3 are determined by Tycho's solar theory; and finally the position of SM is determined by the vicarious theory of Mars. We get three triangles with all angles known, and one side, SM, in common. It is then possible to find the ratios of SE_1, SE_2, SE_3 to SM, and hence to one another; so the positions E_1, E_2, E_3 are found. Three points determine a circle, and so Kepler can determine the position of the center of this circle, and hence the length SC in relation to the radius of the circle.

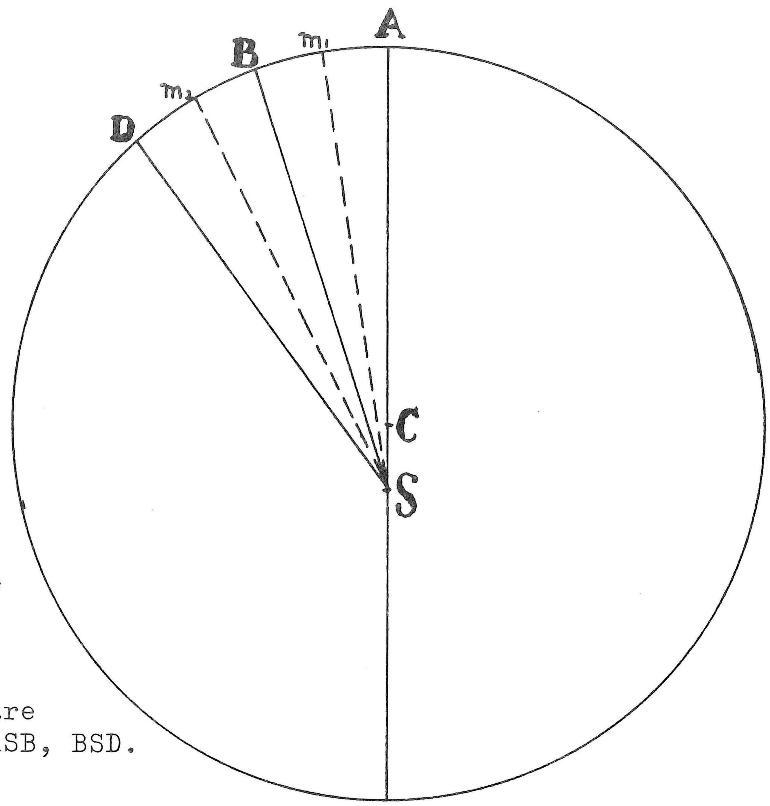
Now I want to stress that these trigonometric calculations are much less satisfactory than the usual accounts of Kepler's work suppose; into each calculation of SC go seven pieces of data, each of which can be in error, and the trigonometric calculation can greatly magnify the initial errors, especially when small angles are involved.

Kepler goes through procedures of this kind several times, and gets divergent results: the highest result for SC is to the lowest as 5 to 3. All of them, however, show that the eccentricity of the orbit is <u>less</u> than the value assigned by Tycho. Remember that the eccentricity determined by Tycho is essentially the eccentricity of the equant, which can always be determined with greater precision than the center of the orbit. Kepler's results, though markedly divergent among themselves, show that the center of the orbit does <u>not</u> coincide with the equant point as had always been previously assumed; rather, it lies somewhere in the middle between the equant point and the Sun. Kepler, then, <u>assumes</u> that the eccentricity is exactly bisected. Then he can proceed to try out his new hypothesis, that the velocities of the planet vary inversely as the distances from the Sun.

The new hypothesis has one great disadvantage: it is difficult to calculate with. The speed of the planet is constantly varying as a function of its distance from the Sun, and the hypothesis says how it is varying; but what we need to know is how far the planet goes in a given time along its path. To determine this, one would need the methods of the calculus, not yet invented. Kepler attempts an approximation: approximately, he says, the times for the planet to traverse equal arcs are proportional to the distances of these arcs from the Sun; the greater the distance, the greater the time (Figure 8). In applying this idea, Kepler goes through very tedious calculations, dividing the semicircle into 180 arcs of one degree each, computing the distancesof each of these arcs from S, adding up successive sums of these distances and putting these sums proportional to the times. Let me mention only one result: in the case of the Earth, predictions based on this new principle differ from those based on Tycho's theory by at most 9" of arc, which is below the level of observational detection. The new theory is satisfactory in predicting the heliocentric longitudes of the Earth, because it jibes with Tycho's solar theory. And Tycho's solar theory, a simple eccentric theory in which the equant point coincides with the center of the orbit, is satisfactory for heliocentric longitudes because the eccentricity of the Earth's orbit is small -- less than 1/5 of the eccentricity of Mars' orbit -- and therefore what matters is only the placement of the equant point, which Tycho has determined fairly precisely. Kepler knows this and depends on it.

FIGURE 8a

STAGE 3

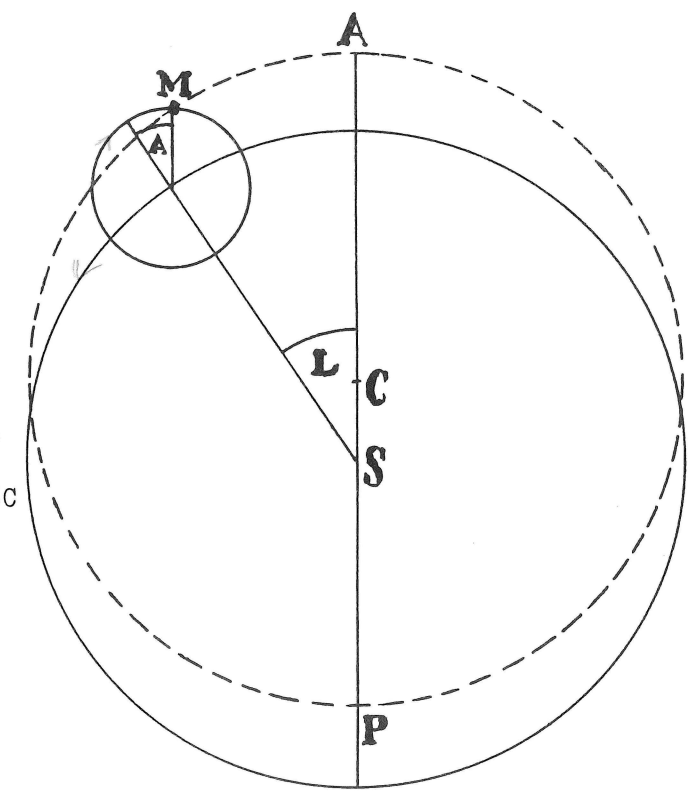


"Radius Rule":
Times for equal arcs AB, BD are as the distances of those arcs from the Sun, Sm_1, Sm_2.

"Area Rule":
Times for equal arcs AB, BD are as the areas of the sectors ASB, BSD.

FIGURE 9

STAGE 4



Ptolemy's solar theory adapted to account for eccentric orbit of Mars

Radius of epicycle = eccentricity SC

S replaces Earth; M replaces Sun

For circular path, angle A must be equal at every moment to angle L

Then Kepler has another idea, to abbreviate the calculations. The distances from the Sun to the points in one of these equal arcs (e.g. AB, Figure 8a) are all contained in the area of the sector, e.g. BSA. The area of this sector suggests itself as a measure of all the distances within the sector. Then just as previously the distances of these equal arcs were proportional to the times, so now the areas might be assumed to be proportional to the times, for these equal arcs. This is the origin of what is called Kepler's second law, the law of areas: as it is usually stated today, the areas swept out by the radius vector from the Sun to the planet are proportional to the times. Actually, this hypothesis is a new principle, not equivalent to the former one. But again, in the case of the Earth, it rhymes satisfactorily with Tycho's theory: the maximum discrepancy is 34" of arc. In the case of Mars, the differences would be much larger.

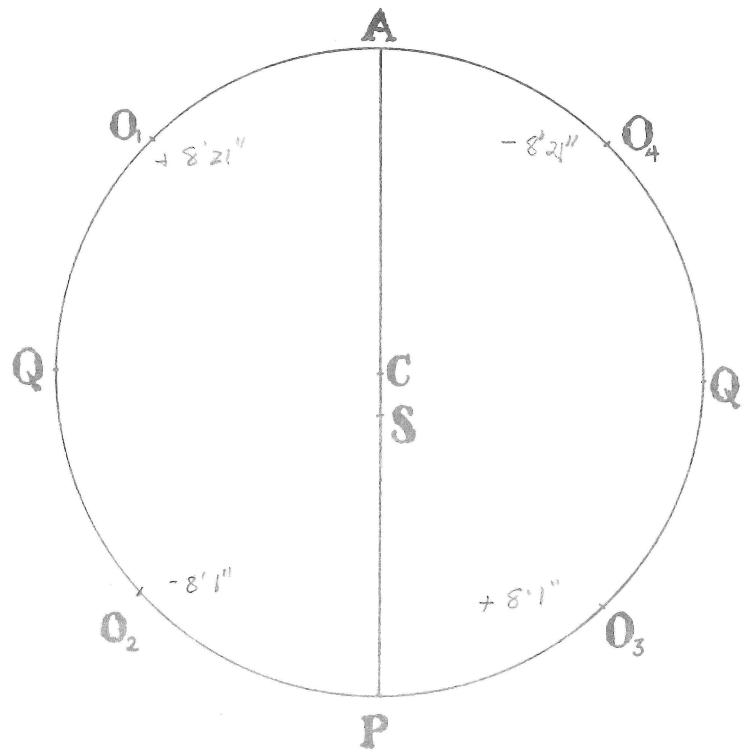
Fourth Stage. Kepler renews his attack on Mars, using his new principle, that the areas are proportional to the times. (I am leaving out of account here the fact that Kepler continued to calculate also with his first hypothesis, that the times for equal arcs are as the distances of these arcs from the Sun. I shall deal only with his use of the area law.)

In this new attack, Kepler first assumes a circular orbit for Mars, with the Sun off-center. The Sun is presumably pushing planet around, the strength of the push decreasing with increasing distance from the sun.

But at this point Kepler has to ask himself a question. Why should the distance of the planet from the Sun vary? Why does it not simply move in a circle concentric with the Sun? The only account Kepler can think of is an adaptation of Ptolemy's epicyclic theory for the Sun (Figure 9). In this, the epicycle has a radius equal to the eccentricity, and moves counterclockwise around a circle with the Earth at the center, while the Sun moves clockwise at the same rate on the epicycle. At every moment, angle L = angle A. On these assumptions, it can be shown that the Sun simply moves in a circle eccentric to the Earth (dotted circle in the figure). Now this same mechanism can be adapted to make Mars move in a circle eccentric to the Sun. Besides being moved around by the Sun, Mars then has a mover of its own, which moves it in this tiny epicyclic circle. Kepler worries about how the planetary mover might accomplish this. He knows,that, as Tycho's observations on the parallaxes of comets have indicated, the heavens are not solid; there are no crystalline spheres; comets move straight through the planetary regions. The planetary mover has no feet or wings. And there is another difficulty. The angle L does not increase uniformly with time, but rather follows the dictates of the area law. But for Mars to remain on a circle, angle A must at each instant be equal to angle L. So the planet has to move non-uniformly about the epicycle. Does the planetary mover have to study planetary tables, trying to find out where it ought to be? Kepler is sceptical about that, but he proceeds anyway to apply the area law of the eccentric circular orbit.

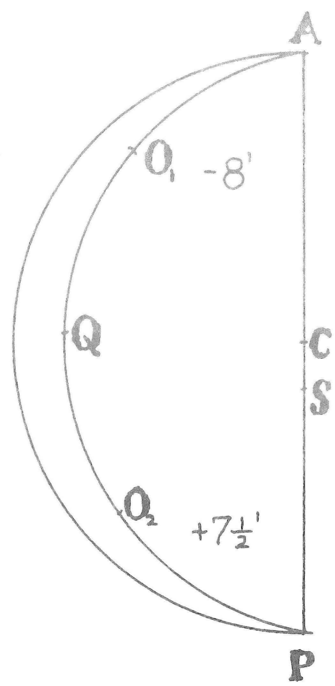
I shall not describe the calculations. Using the area rule, Kepler is finding out where the planet is in its circle at given times in the whole period of 687 days in which it completes its orbit. The results are checked against the predictions of the vicarious theory. As it turns out, there is agreement of

FIGURE 10
STAGE 4



Errors in Circular Theory of Stage 4, as compared with the Vicarious Theory.

FIGURE 11
STAGE 5

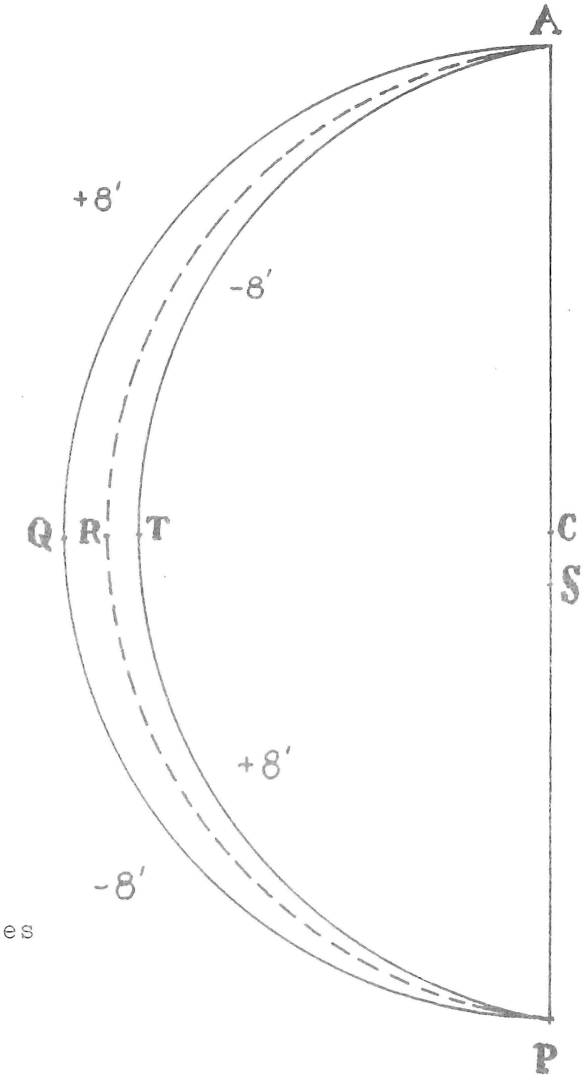


Errors in the Oval Theory of Stage 5, as compared with the Vicarious Theory the two theories in the apsides (A, A) and in the quadrants (Q, Q) but a discrepancy of about 8' in the octants (Figure 10). At O_1 for instance, the new theory puts the planet 8'21" ahead of where it should be; and at O_2, it is 8'1" behind where it should be. The pattern of error is symmetrical in the other semicircle. So the planet, on the assumption of a circular orbit and the area rule or law, is being made to move too rapidly about the apsides, and too slowly about the quadrants. Therefore either the circular orbit must be wrong, or the area rule, or both. If it is only the circle which is wrong, then the orbit must be brought within the circle in the middle longitudes, around the quadrants, so that the amount of area is reduced there, and hence the times for given arcs shortened. The orbit would be oval.

Kepler performs a number of calculations with triangles involving the Earth, Sun and Mars, in order to determine the distance of Mars from the Sun in these middle longitudes. I have previously mentioned the error in this type of calculation. Kepler is now in a better position to make these distance calculations than before, because his modification of Tycho's solar theory -- his introduction of the bisection of the eccentricity -- makes the theory a better predictor of Earth-Sun distances. But the error is still very troublesome. However, Kepler's results do show that Mars comes within the circle in the middle longitudes. In other words, the orbit is some kind of oval, rather than circular. So the circle of the Stage Four theory is wrong. On the other hand, the area law may still be right; it predicts that the orbit is oval, and the distance-determinations, though rough, confirm this prediction.

Fifth Stage. At the beginning of Chapter 45 of the *Astronomia Nova*, Kepler tells us that, having discovered that the orbit is not circular, he felt he knew the cause of the departure from circularity. From Stage Four you will recall that the planetary mover that was moving Mars on its little epicycle was having a rather difficult time. Not only did it lack feet or wings, but it was having to move the planet non-uniformly on the epicycle in order to keep to the eccentric circular path. Its job would be easier, though still impossible, if it had only to move the planet uniformly on the epicycle. But if it were to do this, the resulting path would be oval, as required. This is easy to show, but I shall omit the demonstration. The actual orbit turns out to be egg-shaped, with the sharper end at the peri-helion, or point of closest approach to the Sun.

Kepler proceeds to calculate the consequences of this new hypothesis, assuming as before the area rule. The calculations are horrendous. To simplify matters, Kepler substitutes an ellipse for the egg-shaped oval; the difference in shape is very small. I shall call this ellipse the auxiliary ellipse. Even then the problem remains difficult, and Kepler tries a number of different routes to its solution. The results in which he finally reposes trust (Figure 11) again show agreement with the vicarious theory in the apsides and quadrants, and discrepancies in the octants. In one calculation, for instance, he finds that the new oval theory puts the planet 8' behind where it should be at the first octant, and $7\frac{1}{2}$' ahead of where it should be at the third octant. Note that the errors in the oval hypothesis, on the assumption of the area rule, are almost the exact opposite of those found in the circular hypothesis: the planet is going too slowly about the apsides, and too rapidly in the middle longitudes. If the area law is right, then the orbit should be less narrow: there needs to be more area between the Sun and the orbit in the middle longitudes, in order to show the planet down there, areas being proportional to times.



Errors in Circular and Oval Theories
Circular

	Circular	Oval
1st octant	+8'	-8'
3rd octant	-8'	+8'

FIGURE 12 STAGE 6

Sixth Stage. Kepler now sees that, on the assumption that the area law is right, he will get a theory that jibes with the vicarious theory, in its predictions of heliocentric longitudes, if he chooses an orbit just midway between the circle of Stage Four and the auxiliary ellipse of Stage Five (Figure 12). Rounding off the errors in the octants, we get an anti-symmetrical array (as shown opposite the diagram). The dotted elliptical orbit, just midway between the circle and the auxiliary ellipse, will reduce these symmetrical errors to zero. In effect, the area law is controlling the shape of the orbit. The areas swept out about the Sun are assumed to be proportional to the times; variously shaped orbits distribute the total area of the orbit in different ways; only one shape of orbit will get the planet to the right place at the right time. On the assumption of the area law, the right orbit can differ only negligibly from the dotted ellipse.

Moreover, Kepler is now able to calculate the precise dimensions of the dotted ellipse. Long before, he had calculated the width, QT, of the lunula which the auxiliary ellipse cuts off from the circle. He has only to halve this width to get QR, and hence RC, the semi-minor axis of the ellipse.

Two remarks. As is well known, the correct ellipse is that ellipse which has the Sun at one focus. The term "forcus" -- first introduced into European mathematical literature by Kepler himself, in the Astronomia Pars Optica of 1604 -- is nowhere used in the Astronomia Nova. At the stage of his journey we have now described (Phase #6), Kepler gives no indication as to whether or not he realized that the Sun was located at the focus. The focal properties were not involved in the discovery of the correct ellipse.

Second Remark. What about the possible role of Mars-Sun distances, determined trigonometrically, in the discovery of the correct ellipse? From three of Kepler's letters, written in December, 1604 and January, 1605, it is apparent that the distance-determinations are misleading him: they are giving the wrong value for the amount of ingression of the orbit within the circle. Let the mean distance from the Sun to the Earth be 100,000. Then the mean distance from the Sun to Mars comes out to be 152,350, approximately. The correct, dotted ellipse comes within the circle by about 660 of these parts. Kepler was getting values of 800 or 900 parts. In a passage of a letter written in May of 1605, after the war was over -- it came to an end about Easter time -- Kepler says that the distance determinations generally left him in doubt by about 100 or 200 parts. By contrast, the assumption of the area law, together with the dotted ellipse, leads to predictions which, Kepler tells us, jibed with the vicarious theory "to the nail". The distance-determinations play an essential role, but this role is mainly negative and admonitory. They show that the vicarious theory is wrong in its predictions of Mars-Sun distances, and that Tycho's solar theory is wrong in its predictions of Earth-Sun distances; in both cases, they indicate that the eccentricity is more nearly bisected; and in the case of Mars, they show that the orbit is not a circle but some kind of oval. After the discovery of the correct elliptical orbit of Mars, and after Kepler has, arguing from analogy, altered his theory of the Earth to make the shape of its orbit oval, the distances in these two orbits can be used in predicting positions of Mars as seen from the Earth; and these predictions can be checked against the observations. But Kepler sees these confirmations as confirming his Mars and Earth theories jointly, not the Martian theory by itself. And in any case, the distance determinations do not lead, in the first place, to the idea of the ellipse, or to its exact dimensions: they are too ridden with error to do so.

Seventh and Final Stage. Kepler has the ellipse and the area law -- the first and the second planetary laws that go under his name. He is in despair. He feels that his triumph over Mars is empty. He cannot explain why the planet should go in this particular orbit.

One day he is considering a diagram of the eccentric circle -- not the correct orbit of Mars, but the orbit of Stage Four (Figure 13). This circle, of course, circumscribes the correct elliptical orbit. If we assign 100,000 parts to the radius of the circle, then the ellipse comes within the circle, in the middle longitudes, by 429 parts; Kepler had calculated this number. Now he had been employing this circle constantly, in his calculations of areas in the ellipse; and always in these calculations, a certain triangle played a role: the triangle with the base SC, where S is the Sun, and C is the center of the orbit, and the apex P is on the circle, at such a point that PC is perpendicular to CS. The angle CPS, Kepler knew, was $5^{\circ}18'$. Kepler tells us that quite accidentally he happened on the secant of CPS, that is PS/PC; it was 100429/100000. "It was as though" -- he says -- "I had awakened from sleep, and seen a new light." In the middle longitudes, between the apsides, the lunula is broadest, and in fact its width, 429 is just the excess of PS over PC.

Kepler immediately arrives at the notion that for other places on the circle, e.g., P', the distance of Mars from the Sun should be given, not by SP', but by the perpendicular projection of SP' onto the corresponding diameter of the circle, viz. P'T.

The last part of the story I am barely touching on, because of its complication, but I do want to urge that Kepler's sudden sense of illumination is not totally unintelligible. There are, I think, two reasons for it. One of them has to do with the relation between the area law, and that first hypothesis which Kepler had proposed to replace the equant, the hypothesis that the times for equal arcs are proportional to the distances of those arcs from the Sun. Kepler knew that these two hypotheses were not equivalent for most orbits; but I think that at the moment of illumination, because of a certain geometrical relation, he got the idea that the two hypotheses would be exactly equivalent for the right elliptical orbit. This is not quite right, and Kepler finds a clear and strictly correct formulation only much later, after the Astronomia Nova was finished.

The other reason has to do with the explanation of the ellipticity of the orbit. Kepler now proceeds to replace the motion of the planet on the epicycle, used in Stages Four and Five, by a libration or oscillation of the diameter of the epicycle that goes through the Sun. The possibility of such an oscillation Kepler had thought of long before, when he was having all his difficulties with the conception of the planetary mover, and was trying to imagine a way in which Mars could be caused to move in an orbit eccentric to the Sun. Kepler finally accounts for the oscillation by a kind of magnetic attraction and repulsion. But the point I want to make here is that, in Kepler's earlier study of this oscillation the triangle P'TS is involved.

Thus this moment of illumination is triggered by an accidental observation; but the reason that the accidental observation seems illuminating is that it suggests a solution to two problems on which Kepler had spent long hours, and with the geometry of which he is thoroughly familiar.

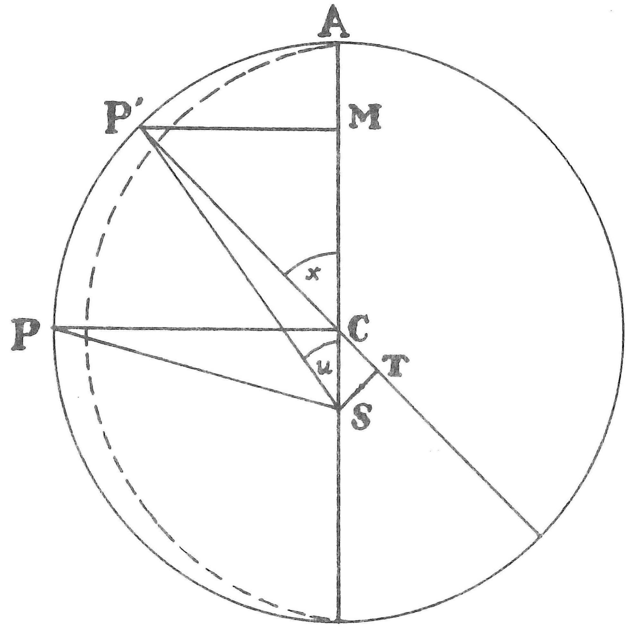


FIGURE 13 STAGE 7

Discovery of the correct rule for Sun-Planet distances

When the angle at the center of the circle is x, and the heliocentric longitude therefore u (approximately), then--

the Sun-Planet distance is not SP' (as in the circular theory),

but P'T (the projection of SP' onto the diameter).

What Kepler now has is a proposed procedure or formula for calculating the distances of Mars from the Sun. For a given angle at the center of the circle, say ∢ x in Figure 13, the proper distance is not SP' but P'T. But: how is P'T to be laid off? One end of it has to go at S, but where does the other end go?

Kepler first verifies his not-quite-defined formula by laying P'T off along SP', the radius vector from Sun to planet in the old, discarded circular theory of Stage Four. This theory, we recall, gave errors in the heliocentric longitudes of 8' in the octants. The resulting orbit, gotten from the new distance formula, applied in this way, is not an ellipse. It does not matter; the angular positions of the new distances are not quite right, but they are never off by more than 8'. In an 8' shift, at the upper octant for instance, the radius vector in the ellipse changes by about 25 parts. The distances determined by observation and trigonometrical calculation are uncertain by 100 or 200 parts. In other words, the observationally determined distances fit orbits which differ slightly in shape from the ellipse.

Kepler knows this. But he wants to fit the distances of the new formula into the <u>ellipse</u>. The orbit <u>must</u> be elliptical, as he has convinced himself on the basis of the area law. After a time-consuming mistake, Kepler discovers the right way. The new distances, laid off from S, are to be shifted in position from the line SP' to a position such that the end-point of the distance lies on P'M, a line perpendicular to the line of apsides. Then the end-points lie on an ellipse. Kepler is here discovering a new piece of geometry, hitherto unknown.

All right. The war is over.

What I have been trying to show may be summarized as follows. The revolution which Kepler brings about in astronomy -- and the Keplerian revolution is the decisive revolution, it is here rather than with Copernicus that we cross the divide between ancient and modern astronomy -- this revolution does not consist in the discovery of what may be called, in a simplistic sense, empirical laws. If the first two Keplerian laws were empirical in that sense, then it would have had to be the case that the ellipticity of the orbit was verified independently of the area law, and then the area law verified within the ellipse thus found. The actual process was the other way round: what Kepler verified, with the degree of precision he wanted, was the proposition: <u>if</u> the area law is right, <u>then</u> the orbit is elliptical. It is true that he has satisfied himself, independently of the area law, on the basis of distance-determinations, that the orbit is oval. But these determinations leave an unsatisfying range of indeterminacy.

Newton will later write: "Kepler knew ye Orb to be not circular but oval, & guest it to be Elliptical." Yes. But the guess is no idle guess; it comes out of a hunch actively pursued, in confrontation with all the previous theories, and with Tycho's new data.

If one is asked what is responsible for Kepler's discoveries, I think one has to admit the role of chance, luck, or as Kepler would say, Providence. It is into his care that Tycho's observations are confided -- the only observations that could have led him to his goal. The first task he is assigned is that of constructing a theory for Mars -- the only planet whose elliptical path could have been discovered, in the then state of the observational art. The theories of Stages Four and Five, which he constructs, happen to err equally in opposite directions from the right result. He happens by accident on the secant of a certain angle, and so emerges from his final perplexity.

But what is just as important, through all the accident and error and luck, is, first, Kepler's belief in the possibility of understanding, and his devotion to his task, that carries him through four years of reasoning and calculation; and secondly, the rightness of his initial hunch or insight, and his ability to disentangle the confused state of things before him in the light of it. He proceeds indeed by a kind of Sherlock Holmesian logic, which claims in eliminating the impossible or false to arrive at the true; and most of the physical hypotheses he constructs will have later to be discarded as inconsistent with Newtonian theory. But at the root of all his theorizing is that initial sense of the significance of the inverse relation between velocity and distance -- a first glimmer of what will one day be the law of conservation of angular momentum. It is in the light of that hunch that he is guided through 900 pages of calculation to a better planetary theory than had ever been proposed before. No doubt Kepler's discoveries are a kind of miracle -- of chance and love, but also of inventive hypothesizing and detective logic.