Why conic sections?

The conic sections are the curves that arise when a cone is cut by a plane. The three fundamentally different kinds of conic sections - ellipse, parabola, and hyperbola - are shown in Figure 1. These terms mean roughly 'too little', 'just right', and 'too much', respectively, which makes sense as characterisations of the inclination of the cutting plane, as we see in the figure. The meanings of these terms are also reflected in the English words ellipsis (the omission from speech or writing of words that are superfluous; also the typographical character '. . .'), parable (a simple story used to illustrate a moral or spiritual lesson) and hyperbole (exaggerated statements or claims not meant to be taken literally).

By **Viktor Blåsjö**

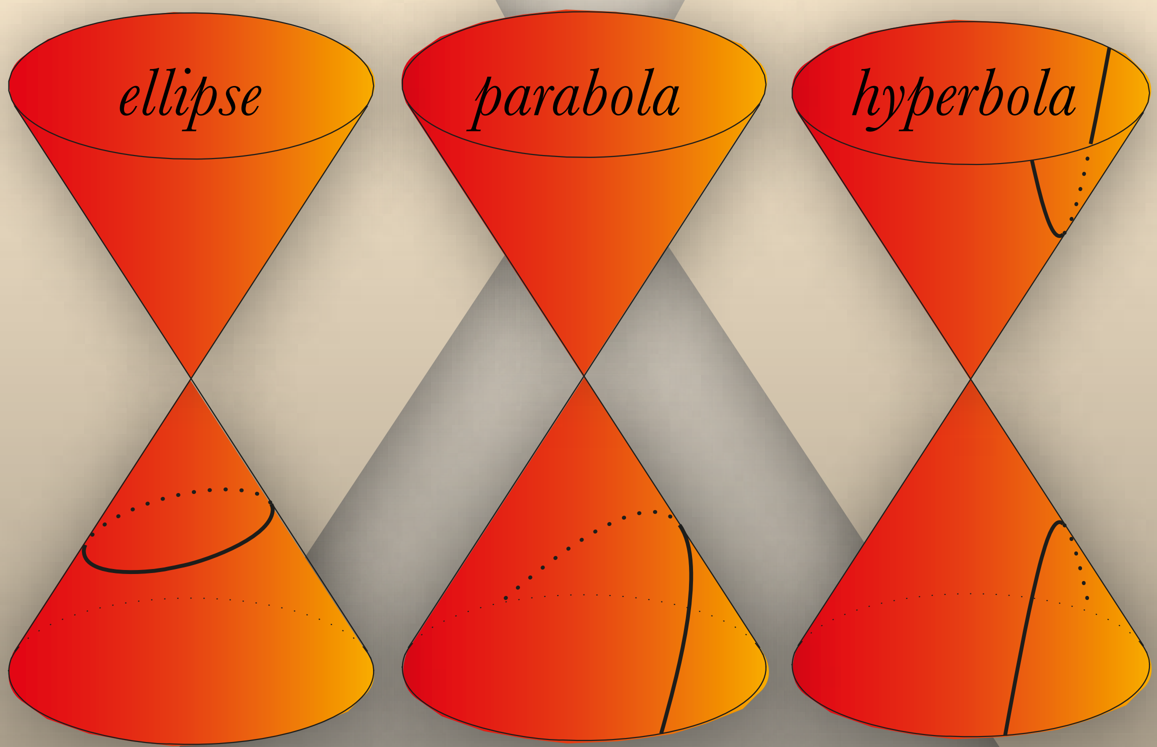


Figure 1: The conic sections. (Figure from [2], p. 29.)

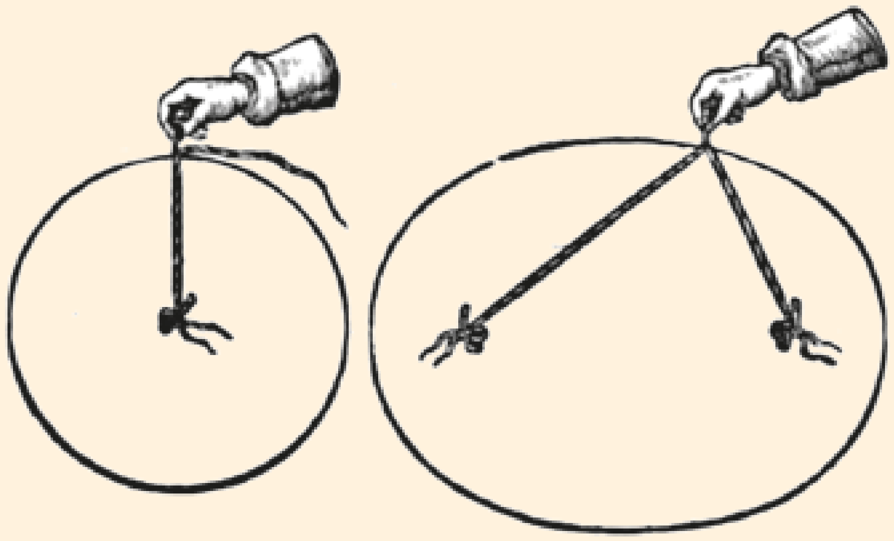


Figure 2: Generating circle and ellipse by means of a string. (Figure from [3], p. 92{93.)

Conic sections were studied in meticulous detail in Greek antiquity well over 2000 years ago. In this article we ask ourselves: Why? What is so interesting about conic sections? Why did the Greeks single them out for such detailed study? There is no straightforward answer. Many roads lead to conic sections and it is not known for certain which was the one history chose first.

Conic sections as 'generalised circles'

One way of looking at it is that conic sections are in many ways the natural next step beyond lines and circles. If we ask ourselves why lines and circles are so fundamental we will find that these reasons can often be generalised to conic sections as well. For example, lines and circles can be generated in very simple ways using a piece of string: stretch the string taunt to make a line, and nail down one of its ends to make a circle. From here it is just a short step to conic sections, for if you nail down both ends of the string you get an ellipse, as shown in Figure 2.

Another way of generating a circle is by a pair of compasses, of course. This construction method too generalises to conic sections in the manner shown in Figure 3. We have no written record of such generalised compasses until the medieval Islamic commentary literature, but it could very well have been known in Greek times. In fact, in a sense it may even be considered implicit in the very definition of a conic section. For what is a cone anyway but a line rotated about an axis? So in this sense the generalised compasses is really nothing but the physical manifestation of the literal meaning of the definition of a conic section. We must note, however, that in the crowning Greek work on conics, that of Apollonius, the definition of a cone includes skew cones as well, for which the generalised compasses do not apply. But this is quite plausibly a late abstraction; unfortunately very little has survived of Greek work on conics before the magisterial treatise of Apollonius, written at a point when the theory had already been developed for a long time and become highly refined.

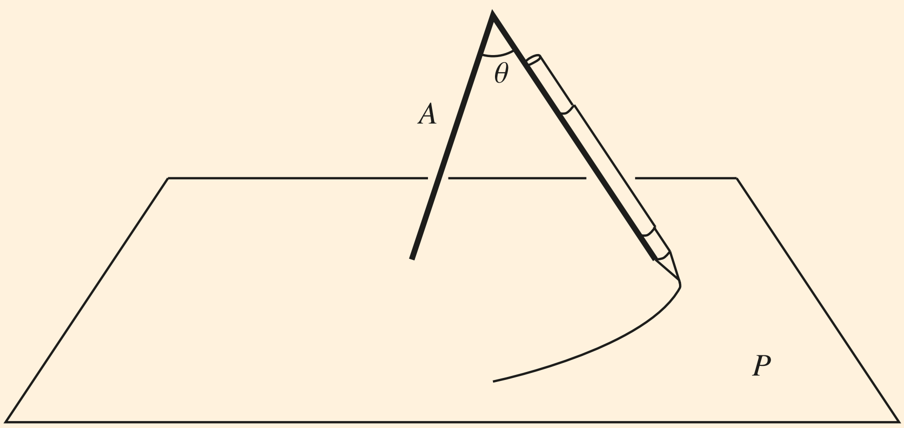
The conic sections are also the natural next step beyond lines and circles in that they are curves of degree 2. Today we are inclined to think of this in algebraic terms: conic sections correspond to algebraic equations of degree 2. The description of curves by algebraic equations is a 17th century invention, however, this was not how the Greeks looked at it. But the concept of the degree of a curve can also be defined in purely geometrical terms, as the greatest number of intersections the curve can have with a line. In this form the degree argument for conic sections was certainly available to the Greeks.

Conic sections as arising from applications

An astronomical application of conic sections was also important in Greek times, namely the fact that the tip of a gnomon (a fancy word for a stick standing in the ground) traces a conic path in the course of a day. This is so since the sun travels in an essentially circular path through the sky. Thus the circular path of the sun together with the point-tip of the gnomon define a cone, and the plane of the ground is a plane cutting this cone, thus producing a conic section.

However, some indications from the early history of conic sections suggest that none of the above reasons were the historically central ones. Very little is known about the early history of conic sections, but arguably the two main facts known about it are: (i) At an early stage conics were used for the duplication of the cube, i.e., the problem of constructing a cube with twice the volume of a given cube. This is one of the most basic questions one can ask regarding volume, so it was a mathematically fundamental problem. It amounts to solving $x^3 = 2$, so it can be accomplished by combining

Figure 3: Generalised compasses for drawing conic sections. The angle θ and the direction of the axis A are fixed. As the other leg rotates around the axis, the pen slides up and down in its cylinder, so as to always reach the plane P. (Figure from [2], p. 30.)



the hyperbola $xy = 2$ with the parabola $y = x^2$. (ii) In the early period, cones were defined as line segments rotated about an axis and conic sections as the intersection of a cone with a plane perpendicular to its side. The perpendicularity restriction in (ii) at first appears very artificial and strange. It makes no sense in terms of the natural application of conic section theory in astronomical gnomonics, nor does it make any theorems about conics easier to prove. This suggests that the study of conic sections was not originally an end in itself, but only a way of interpreting curves already necessitated elsewhere. The solution of (i) came first, and the notion of a conic section was concocted as a way of explicating the curves involved in this important construction. From this point of view the perpendicularity restriction is reasonable, since it leads to definite and clear constructions of all conic curves; greater generality is superfluous and would only muddle the matter. (This hypothesis regarding the origins of conic section theory was proposed by Zeuthen [4], chapter 21.)

Astronomy after all?

Neugebauer [1] proposed that the theory of conic sections originated in gnomonic astronomy after all, and that in fact the perpendicularity condition in the early definition of conics is evidence not against this hypothesis but in favour of it. Of course this condition does not make sense in terms of a perpendicular gnomon casting a shadow on the ground. But consider a different kind of sundial, where the gnomon is pointing towards the highest (i.e., noon) position of the sun in any given day, and the plane recording its shadow is perpendicular to this gnomon. This arrangement, sure enough, produces conic sections consistent with the perpendicularity condition, for in this case the gnomon is contained within the surface of the cone defined by the circular path of the sun and the tip of the gnomon, whence the plane perpendicular to the gnomon is also perpendicular to the side of the cone, as required.
Neugebauer declared himself “confident that the above explanation gives the real motivation for the early Greek theory of conic sections,” but I find his confidence difficult to share. The obvious problem with this hypothesis is that there is no evidence that such sundials were ever considered in ancient times. It is true that Neugebauer’s sundial has certain theoretically pleasing properties, as he notes with the insinuation that this ought to have been appreciated by ancient astronomers. But these pleasing properties, it seems to me, do not come from this sundial being more natural or superior but simply from removing much of the complexity of a traditional sundial by the mere stipulation that the gnomon is always pointing toward the highest point of the sun. Certainly, if this stipulation was workable in practice this sundial would have much to commend it. But it would be very cumbersome in practice to keep realigning the direction of the sundial every day so as to keep up with the sun’s changing position across the seasons. If anything, the easiest way to determine the highest point of the sun is by means of a fixed sundial, so Neugebauer’s sundial has a distinct note of circularity in its setup. It is a mathematician’s fantasy more than a realistic proposal.
We are left with the conclusion that although conic sections are natural from many points of view, it seems that they entered mathematics not by any of these obvious, natural paths but only indirectly, as subservient tools for addressing a technical problem. Astonishing as it may seem in retrospect, many great mathematical concepts were in fact originally introduced in such unglamorous fashion. !

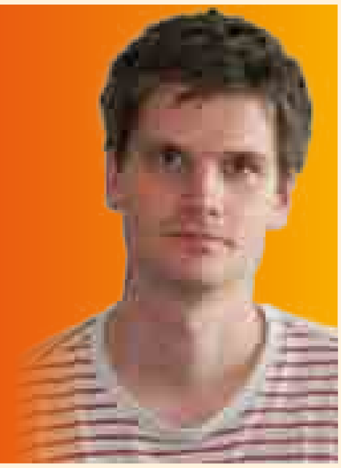
Certainly, if this stipulation was workable in practice this sundial would have much to commend it.

References

[1] Neugebauer, Otto. The Astronomical Origin of the Theory of Conic Sections. Proceedings of the American Philosophical Society, 92(3) (1948), 136138.
[2] Stillwell, John. Mathematics and Its History. 3rd ed. Springer, 2010.
[3] Tschirnhaus, EhrenfriedWalther von. Medicina mentis sive artis inveniendi praecepta generalia. Leipzig, 1695.
[4] Zeuthen, H. G. Die Lehre von den Kegelschnitten im Altertum. A. F. Höst & Sohn, 1886.

About the author - Viktor Blåsjö

Viktor Blåsjö is a Ph.D. student in history of mathematics at Utrecht University. He is originally from Sweden and came to Utrecht because of its unique tradition of research in the history of mathematics. His thesis subject is the representation of transcendental curves in the late 17th century, which is a continuation of a research program initiated by Utrecht professor emeritus Henk Bos.



www.uu.nl/staff/VNEBlasjo1
V.N.E.Blasjo@uu.nl