History of Mathematics

Viktor Blåsjö

Contents

§ 0. What is history?

Consider this quotation from Isaac Newton, in which I have left out the subject of the sentence:

> $\boxed{}$ is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it.

What do you think he is talking about? Astrology perhaps, or some such pseudoscience? No, what goes in the blank is "that one body may act upon another at a distance through a vacuum without the mediation of any thing else." In other words, the "great absurdity" that Newton is referring to is his own theory of gravity. And isn't he right? Isn't it absurd that one planet can pull at another through thousands of miles of empty space? And that as I move my hand I move the moon? Is this any crazier than alchemy, astrology, or witchcraft? Hardly.

In the history of science, things that make sense happen before things that don't. That is why Newton's absurd theory didn't see the light of day until 1687, thousands of years after people had started speculating about the heavens along the much more natural lines of numerology and astrology. The question we ask ourselves as historians is not "when did people get it right?" but rather "why did people do it this other way, and why did it make sense to them?"

Modern schoolteachers demand that you disregard thousands of years of common sense and swallow Newton as a dogma, and, more generally, that you embrace anything you are told regardless of whether it serves any credible purpose for you at that time. If you were ever dissatisfied with this state of affairs then history is on your side. What Alfred North Whitehead said of education is in effect a description of history:

> Whatever interest attaches to your subject-matter must be evoked here and now; whatever powers you are strengthening in the pupil, must be exercised here and now; whatever possibilities of mental life your teaching should impart, must be exhibited here and now. That is the golden rule of education, and a very difficult rule to follow.

History works this way because it cannot "look ahead" and see what will become useful later, as the curriculum planner does. Historically, ideas occur when they serve a purpose, and not a day sooner. Thus we see what Ernst Mach meant when he wrote:

> The historical investigation of the development of a science is most needful, lest the principles treasured up in it become a system of half-understood prescripts, or worse, a system of prejudices.

Or as Descartes put it:

> To converse with those of other centuries is almost the same thing as to travel. It is good to know something of the customs of different peoples in order to judge more sanely of our own, and not to think that everything of a fashion not ours is absurd and contrary to reason, as do those who have seen nothing.

Or Poincaré:

> By going very far away in space or very far away in time, we may find our usual rules entirely overturned, and these grand overturnings aid us the better to see or the better to understand the little changes which may happen nearer to us, in the little corner of the world where we are called to live and act. We shall better know this corner for having traveled in distant countries with which we have nothing to do.

§ 1. Astrology

Where does mathematical and scientific inquiry begin? Imagine a primitive man standing in the middle of a field. Looking at the world around him, what will spark his interest and give him reason for reflection? Will he look at a falling apple and ask himself what equations describe its velocity and acceleration? No, why would he? He cannot fail to notice, however, that half the time the sky goes black and a beautiful spectacle of sparkling lights is displayed as if only for him. Surely it would be an insult to the creator of the universe not to observe this play written in the sky.

The idea soon suggests itself that the heavens control earthly affairs. They obviously determine day and night, and the seasons, and the connection between the moon and tidal waters will be unmistakable to anyone living in a costal area. Likewise the female menstrual cycle follows the moon's periods. The stars too play their role: in the days before calendars people used them to tell the time of the year. Thus one finds in ancient texts such sound advice as "when strong Orion begins to set, then remember to plough" and "fifty days after the solstice is the right time for men to go sailing."

Once you start thinking like this you will soon begin to see the periodicity of the heavens mirrored in the cycle of life, the rise and fall of empires, and so on. Heavens are also the paradigm of self-motion. Most motion is derived from them: we already mentioned the tides, and likewise wind, rain, and rivers are all caused by differential heating from the sun. In fact, the only things that move out of their own power are heavenly bodies and things that have a soul. Is it not a straightforward conclusion, then, that the soul is a piece of "stardust" inhabiting an earthly form? The fundamental tenet of personal astrology is a most natural corollary: the soul, on its way down through the planetary spheres at the moment of birth, acquires its particular character depending on the positions of the planets at that moment.

1.1. Criticise the arguments against astrology put forth by Carl Sagan in the readings.

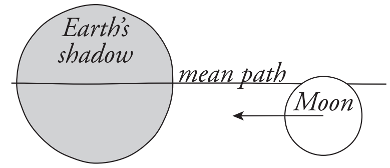
Sagan's remarks embody a common attitude toward historical thought: closed-minded judgement in light of modern views. This is precisely the kind of attitude that we must leave behind if we hope to ever understand anything of history. I chose to start my lectures with the subject of astrology since it is a great subject for bringing this issue to the fore in an emphatic way.

Astrology made a lot of sense once upon a time and was still practiced in earnest by some of the best scientists of the early 17$^{\text{th}}$ century. Kepler did lots of astrology, of which we shall see a few indications later. "The belief in the effect of the constellations derives in the first place from experience, which is so convincing that it can be denied only by people who have not examined it," he said. Galileo also did astrology, and evidently not only to make money judging by the fact that he cast horoscopes for his own children. Soon thereafter astrology fell out of favour, probably largely due to the rise of the mechanical philosophy that we shall hear more about later.

1.2. Explain how the ancient conceptions of the "personalities" of the heavenly bodies (quoted in the Beck reading) all have a certain amount of rationale in terms of the astronomical properties of these bodies.

Much sophisticated mathematics originated in a context such as this. For example, van der Waerden suggested in his book *Geometry and Algebra in Ancient Civilizations*, p. 32, that the original motivation for the discovery of the Pythagorean Theorem might have been the following problem.

1.3. Find the duration of a lunar eclipse as a function of the moon's deviation from its mean path. You need to use the facts that the moon's speed is known and that the earth's shadow as cast on the moon has about twice the radius of the moon.



Today we teach the Pythagorean Theorem as if it had applications to measuring lengths of various kinds. But why would anyone use this theorem to measure the distances between between various locations or the sizes of fields, etc.? In most such cases you can just measure the sought length directly, so no one would ever have any need for the theorem even if he knew it, let alone have reason to discover it in the first place.

1.4. Discuss some other standard applications of Pythagoras' Theorem and whether they serve any practical purpose that cannot equally well be accomplished by direct measurement and simple trial-and-error methods.

The eclipse problem is very different. Here there is no possibility of a direct measurement. In fact, in this context there is no possibility of discovering Pythagoras' Theorem by trial and error (e.g., based on noticing regularities in measurements of lots of triangles). So in fact it suggests a reason to discover not only the theorem but also its *proof.* Perhaps something along those lines is how deductive mathematics began.

1.5. The ancient Babylonians didn't have telescopes or calculators or Wikipedia or anything beyond middle school mathematics, and yet they were able to determine the lunar period (the average time between one new moon and the next) to an accuracy of within *one second.* How did they do this? One trick would be to measure the time between full moons many years apart and divide by the number of full moons that have passed in between. It is difficult, however, to pinpoint the precise moment when the moon is exactly full.

   (a) Suppose you can pinpoint the moment of full moon with an error of at most 12 hours. Then how many years of observation are needed to ensure a final accuracy of one second?

   Instead of counting the time from full moon to full moon you might count from lunar eclipse to lunar eclipse.

   (b) Would there still be an integer number of months between two lunar eclipses?

   A lunar eclipse is easier to pinpoint in time since it is a very distinctive occurrence that only lasts for a short while.

(c) Suppose you can pinpoint the moment of full moon with an error of at most half an hour. Then how many years of observation are needed to ensure the final accuracy of one second?

§ 2. Numerology

Another thing you notice when studying the sky is that there are precisely seven heavenly bodies moving with respect to the stars: the sun, the moon, Mercury, Venus, Mars, Jupiter, Saturn. (Figure 1.) So Nature has chosen the number seven. This made a great impression on early man, so much so that he put sevens in all kinds of places, such as the seven days of creation (and hence in a week) and the seven notes in a musical scale.

Figure 1: Astronomers studying the seven heavenly bodies.

"All is number," said the Pythagoreans, and they had good reason to. For they discovered that musical harmony is determined by simple integer ratios such as 2:1, 3:2, 4:3, etc. See figure 2. If you put a $\sqrt{2}$ bell in there it sounds like crap.

Figure 2: Pythagorean discovery that musical harmony is determined by integer relationships.

Later Newton put seven colours in the rainbow because he imagined a parallel with music in which pleasant colour combinations are harmonious "visual chords," so to speak.

Another special number is five, because there are precisely five regular polyhedra (figure 3).

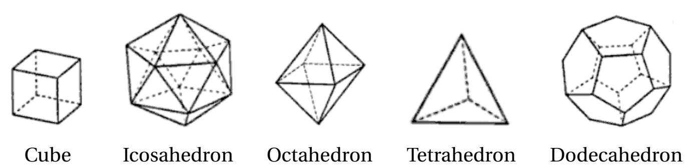


Figure 3: The five regular polyhedra.

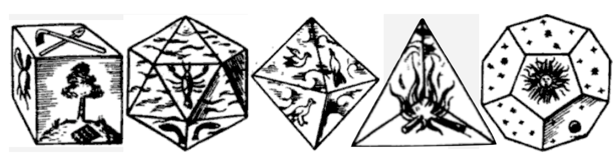


Figure 4: Regular polyhedra as corresponding to the elements: earth, water, air, fire, and "the universe."

2.1. Prove this by enumeration of cases (as Euclid does in the last book of the *Elements*).

Plato proposed that the regular polyhedra correspond to the elements, as we see in the readings. Many centuries later Kepler was enticed by this theory and drew the illustration in figure 4.

The two special numbers five and seven are built into man, for we have five fingers, five senses (sight, touch, taste, smell, hearing), and seven "windows of the head" (eye, eye, ear, ear, nose, nose, mouth). The Creator must truly consider us the crown of his achievement.

The Greeks often thought of the factors of a number as its "parts." Thus for example the number 4 represented justice since it is the smallest number made up of two equals, $4 = 2 \times 2$. The number 7 is special from this point of view also, as Aristotle explains (fragment 203):

> Since the number 7 neither generates [in the sense of multiplication] nor is generated by any of the numbers in the decad [i.e., the first ten numbers], they identified it with Athene. For the number 2 generates 4, 3 generates 9, and 6, 4 generates 8, and 5 generates 10, and 4, 6, 8, 9, and 10 are also themselves generated, but 7 neither generates any number nor is generated from any; and so too Athene was motherless and ever-virgin.

When the factors of a number are considered its parts it becomes natural to ask whether all numbers are the sum of its parts. In fact this is not so; very few numbers are "perfect" enough to have this pleasant property, as Nicomachus (c. 100) explains:

> When a number, comparing with itself the sum and combination of all the factors whose presence it will admit, it neither exceeds them in multitude nor is exceeded by them, then such a number is properly said to be perfect, as one which is equal to its own parts. Such numbers are 6 and 28; for 6 has the factors 3, 2, and 1, and these added together make 6 and are equal to the original number, and neither more nor less. 28 has the factors 14, 7, 4, 2, and 1; these added together make 28, and so neither are the parts greater than the whole nor the whole greater than the parts, but their comparison is in equality, which is the peculiar quality of the perfect number.
>
> It comes about that even as fair and excellent things are few and easily enumerated, while ugly and evil ones are widespread, so also are the superabundant and deficient numbers found in great multitude and irregularly placed, but the perfect numbers are easily enumerated and arranged with suitable order; for only one is found among the units, 6, only one among the tens, 28, and a third in the ranks of the hundreds, $\boxed{\quad\quad}$, and a fourth within the limits of the thousands, 8128.

Euclid proved that if p is a prime and $2^p - 1$ is also prime then $2^{p-1}(2^p - 1)$ is perfect. This is the grand finale of Euclid's number theory (*Elements* IX.36). The theorem amounts to a recipe for finding perfect numbers: in a column list the prime numbers; in a second column the values $2^p - 1$; cross out all rows in which the second column is not a prime number; for the remaining rows, place $2^{p-1}(2^p - 1)$ in the third column. Then the numbers in the third column are perfect numbers.

2.2. Find the perfect number omitted in the Nicomachus quote above using Euclid's recipe. What prime p did you need to use?

The following is essentially Euclid's proof of the theorem. If $2^p - 1$ is prime, it is clear that the proper divisors of $2^{p-1}(2^p - 1)$ are $1, 2, 2^2, \ldots, 2^{p-1}$ and $(2^p - 1), 2(2^p - 1), 2^2(2^p - 1), \ldots, 2^{p-2}(2^p - 1)$. So these are the numbers we need to add up to see if their sum equals the number itself.

2.3. (a) Show that $1 + 2 + 2^2 + \ldots + 2^{p-1} = 2^p - 1$ by adding 1 at the very left and gradually simplify the series from that end.

(b) Use a similar trick for the remaining sum, and thus conclude the proof.

§ 3. Origins of geometry

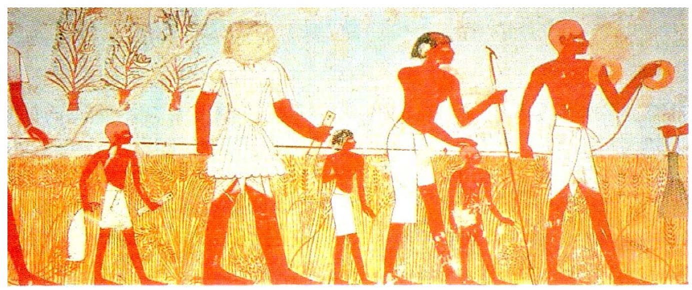


Figure 5: Egyptian geometers, or "rope-stretchers" as they were called, delineating a field by means of a stretched rope.

"Geometry" means "earth-measurement," and indeed the subject began as such, according to ancient sources such as Proclus and Herodotus, as we see in the readings. This was necessitated by the yearly overflowing of the Nile in Egypt: the flooding made the banks of the river fertile in an otherwise desert land, but it also wiped away boundaries between plots, so a geometer, or "earth-measurer," had to be called upon to redraw a fair division of the precious farmable land. In fact the division was perhaps not always so fair, as Proclus also suggests, for one can fool those not knowledgable in mathematics into accepting a smaller plot by letting them believe that the value of a plot is determined by the number of paces around it.

3.1. Prove that a square has greater area than any rectangle of the same perimeter.

3.2. Discuss what general point about history we can learn from the following paraphrase of Proclus's remark in Heath's *History of Greek Mathematics* (1921): "[Proclus] mentions also certain members of communistic societies in his own time who cheated their fellow-members by giving them land of greater perimeter but less area than the plots which they took themselves, so that, while they got a reputation for greater honesty, they in fact took more than their share of the produce." (206–207)

Among the first things one would discover in such a practical context would be how to draw straight lines and circles. In fact you need nothing but a piece of string to do this.

3.3. Explain how.

3.4. Problem 3.1 shows that it is important to be able to construct squares. How would do this with your piece of string?

3.5. In the Rhind Papyrys (c. −1650) the area of a circular field is calculated as follows: “Example of a round field of a diameter 9 khet. What is its area? Take away $\frac{1}{9}$ of the diameter, namely 1; the remainder is 8. Multiply 8 times 8; it makes 64. Therefore it contains 64 setat of land.” What is the value of π according to the Rhind Papyrys?

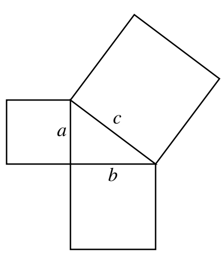
People soon recognised the austere beauty of geometrical constructions and began using it for decorative and especially religious purposes. Indeed, Egyptian temples are very geometrical in their design; the famous pyramids are but the most notable cases. One of the first decorative shapes one discovers how to draw when playing around with a piece of string is the regular hexagon.

3.6. Show how this is done.

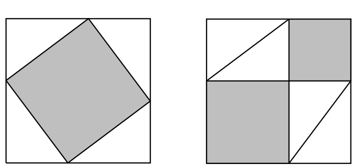
The hexagon has great decorative potential since it can be used to tile the plane. Hexagonal tiling patterns occur in Mesopotamian mosaics from as early as about -700.

3.7. Show that the hexagon contains even more area than a square of the same perimeter. As Pappus explains in the readings, bees seem to know this.

The step from this kind of decorative and ritualistic pattern-making to deductive geometry need not be very great. In fact, two of the most ancient theorems of geometry could quite plausibly have been discovered is such a context. Take for instance the Pythagorean Theorem. Its algebraic form “$a^2 + b^2 = c^2$” seems to be the only thing some people remember from school mathematics, but classically speaking the theorem is not about some letters in a formula but actual squares:

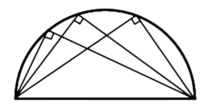


The simplest case of the theorem, when the two legs are equal ($a = b$), is very easy to see when looking at a tiled floor, as we see in the reading from Plato’s *Meno*. Inspired by this striking result, ancient man might have gone on to consider the case of a slanted square, and then discovered that with some easy puzzling the theorem is easily generalised to this case as well:



3.8. Explain how this proves the theorem.

The Greek tradition has it that Thales (c. −600) was the first to introduce deductive reasoning in geometry. One of the theorems he supposedly dealt with was “Thales’ Theorem” that the triangles raised on the diameter of a circle all have a right angle:



3.9. Explain how Thales’ Theorem can very easily be discovered when playing around with making rectangles and circles. Hint: Construct a rectangle; draw its diagonals; draw the circumscribed circle.

Thus we see a fairly plausible train of thought leading from the birth of geometry in practical necessity, to an appreciation for its artistic potential, to the discovery of the notions of theorem and proof.

Another indication of the use of constructions is the engineering problem of digging a tunnel through a mountain. Digging through a mountain with manual labour is of course very time-consuming. It is therefore desirable to dig from both ends simultaneously. But how can we make sure that the diggers starting at either end meet in the middle instead of digging past each other and making two tunnels?

3.10. Solve this problem using a rope. (You may assume that the land is flat except for the mountain.)

Such methods were used in ancient times. On the Greek island of Samos, for example, a tunnel over one kilometer in length was dug around year −530, for the purpose of transporting fresh water to the capital. It was indeed dug from both ends.

§ 4. Babylonia

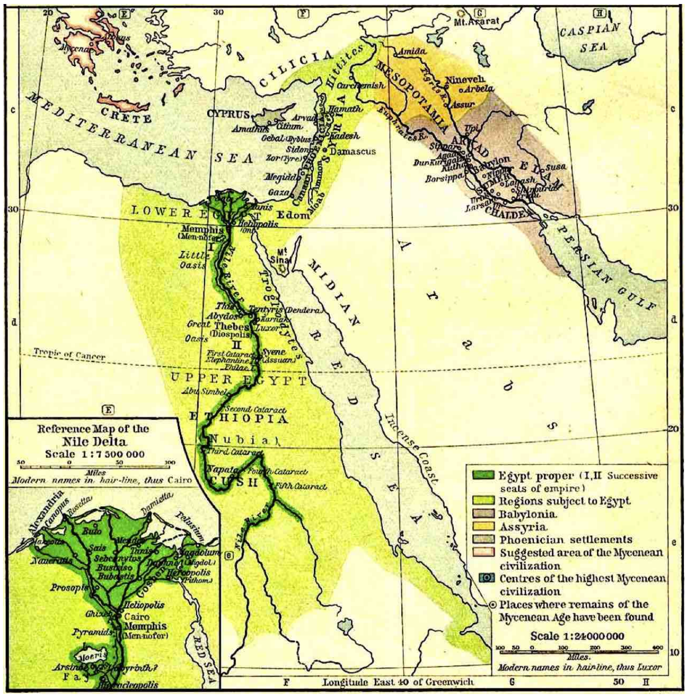


Figure 6: Early mathematical activity coincides with favourable agricultural conditions.

Mathematics arose in the Nile area since the river made the soil fertile and allowed sufficiently rich harvests for some people to concern themselves with intellectual pursuits going beyond daily needs. The same conditions produced mathematics in ancient Babylonia, and indeed their mathematics too can be seen as arising quite naturally from land-surveying. This is illustrated in the following two examples, taken from clay tablets written around −1800, give or take a century or two.

The tablet in figure 7 has to do with the diagonal of a square. The numbers are given in sexagesimal (base 60) form, a Babylonian invention that still lives in our systems for measuring time and angles. Why did they use base 60?

4.1. Argue that 60 has favourable divisibility properties. In which context might this have been important? Hint: in which contexts do people count in "dozens"?

4.2. Explain how you can count to 60 on your fingers in a natural way. Hint: curl your fingers.

4.3. Argue that in the reader there are passages that can be seen as supporting each of these factors as explanations for the origin of the base-60 system.

Base 60 means that, for example, $42,25,35 = 42 + \frac{25}{60} + \frac{35}{60^2}$.

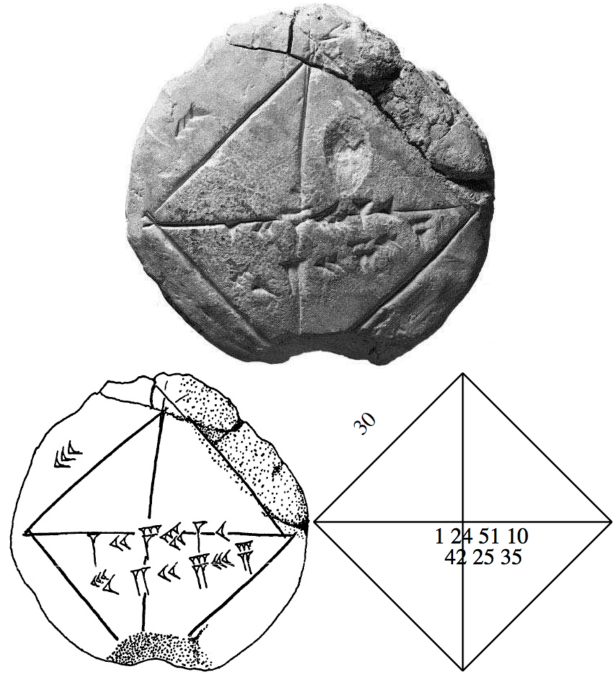


Figure 7: Clay tablet YBC 7289 from the Yale Babylonian Collection.

4.4. (a) Explain the meaning of the numbers on the tablet above. Hint: there are three numbers and one of them is $\approx \sqrt{2}$.

(b) Convert the tablet's value for $\sqrt{2}$ into decimal form.

(c) If you use this value to compute the diagonal of a square of side 100 meters (i.e., roughly the size of a football field), how big is the error? Draw this length.

The Babylonians were very good at solving problems that in our terms correspond to quadratic equations. Such problems are related to areas of fields, though the problems solved on the tablets ask contrived questions that go beyond any practical need and seem to serve no other purpose than posing challenges or showing off one's skills. So one can easily imagine that this mathematical tradition stemmed from practical land-measurements which eventually produced a specialised class of experts who started taking an interest in mathematics for its own sake.

The following is an example of such a problem. I give here the translation of Høyrup; in the reading from his book you will find some further discussion of its context and significance.

"The surface and my confrontation I have accumulated: 45' it is." It is to be understood that "the surface" means the area of a square, and the "confrontation" its side. So the problem is $x^2 + x$=45'. Again the numbers are sexagesimal, so 45' means 45/60=3/4.

"1, the projection, you posit." This step gives a concrete geometrical interpretation of the expression $x^2 + x$. We draw a square and suppose its side to be x. Then we make a rectangle of base 1 protrude from one of its sides. This rectangle has the area $1 \cdot x$, so the whole figure has the area $x^2 + x$, which is the quantity known.

"The moiety of 1 you break, 30' and 30' you make hold." We break the rectangle in half and attach the half we cut off to an adjacent side of the square. We have now turned our area of 45' into an L-shaped figure.

"15' to 45' you append: 1." We fill in the hole in the L. This hole is a square of side 30', so its area is 15'. So when we fill in the hole the total area is 45'+15'=1.

"1 is equalside." The side of the big square is 1.

"30' which you have made hold in the inside of 1 you tear out: 30' is the confrontation." The side of the big square is x+30' by construction, and we have just seen that it is also 1. Therefore x must be 1-30'=30', and we have solved the problem.

4.5. Illustrate the steps of the solution with figures.

4.6. Give a modern algebraic solution to $x^2 + x = 3/4$ which corresponds to the above step by step.

§ 5. Euclid

Whatever the beginnings of Greek geometry, the written record available to us only begins with Euclid's *Elements* (c. −300). As mathematicians we adore it as a truly masterful synthesis of the mathematical canon, but as historians we also regret that its polished refinement no doubt erases many intricacies of the three centuries of development following Thales.

Euclid's *Elements* was the paradigm of mathematics for millennia. It embodies the "axiomatic-deductive" method, i.e., the method of starting with a small number of explicitly stated (and preferably obvious) assumptions and definitions and then deducing everything from there in a logical fashion, thereby compelling anyone who believed in the original assumptions to also believe in all subsequent theorems. Some selections from the definitions of Euclid's *Elements* are the following:

> Definition 1. A *point* is that which has no part.
>
> Definition 2. A *line* is breadthless length.
>
> Definition 4. A *straight line* is a line which lies evenly with the points on itself.
>
> Definition 10. When a straight line standing on a straight line makes the adjacent angles equal to one another, each of the equal angles is *right*.
>
> Definition 15. A *circle* is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure equal one another.
>
> Definition 23. *Parallel* straight lines are straight lines which, being in the same plane and being produced indefinitely in both directions, do not meet one another in either direction.

The following are all the geometrical assumptions admitted by Euclid:

> Postulate 1. To draw a straight line from any point to any point.
>
> Postulate 2. To produce a finite straight line continuously in a straight line.
>
> Postulate 3. To describe a circle with any center and radius.
>
> Postulate 4. That all right angles equal one another.
>
> Postulate 5. That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

The first two postulates can be seen as saying roughly that one has a ruler (or rather an unmarked ruler, or "straightedge"). Similarly the third postulate basically grants the existence of compasses. In this sense we are still not so far from the banks of the Nile, as it were. And in fact this is true also geographically, for Euclid's *Elements* was written in Alexandria.

The fifth postulate is the most profound one, and was the source of much puzzlement for thousands of years following Euclid. We shall now look at some of Euclid's theorems in which this postulate is crucial in the hopes of better understanding both this postulate and the axiomatic-deductive method generally.

Euclid's very first proposition is the construction of an equilateral triangle from a given line segment.

5.1. (a) Explain how to do this with ruler and compasses.

(b) Find a hidden assumption in this argument that is not supported by the postulates and definitions.

Suppose you wanted to construct a line parallel to a given line through a given point. It may seem that this would be very easy by essentially just constructing two equilateral triangles next to each other.

5.2. Explain how you might do this.

So you might say: since this is such a simple extension of Euclid's first proposition, he might as well have placed it as his second proposition. But the matter is not so simple. You need to prove that the line you constructed is parallel to the given line using nothing but Euclid's definitions and postulates. Perhaps you found it "obvious" the line you constructed was parallel to the first, but the whole point of the Euclidean way of doing things is to never take anything for granted except what you can strictly infer from the axioms.

5.3. When we feel that the above construction "obviously" gives a parallel, we may, strictly speaking, be thinking of a different notion of parallel than the one of Euclid's definition, namely what alternative definition of parallels? Of course in a stringent logical treatise you have to pick a definition and stick with it; ambiguity has no place in such a work.

It is clear, then, that any proof about parallels must ultimately trace back to the parallel postulate. But this postulate says: To check if two lines are parallel, draw a third line across and add up the two "interior" angles it makes with these lines on, say, its right hand side. If those angles are less than two right angles (180°) then the lines will meet at that side. It follows that if the angles are greater than two right angles then the lines will meet on the other side. Note that the postulate doesn't say when lines *are* parallel, only when they are *not*. The postulate rules out any situation when the sum of the angles is not two right angles, so for parallels to exist at all this must happen in the case when the sum of the angles is precisely two right angles. Indeed this is precisely what happens, of course, but Euclid hasn't put this in the postulate because he can prove it as a theorem (proposition 27). We shall not follow Euclid's specific approach in detail, but the following line of reasoning captures his spirit and will help us appreciate the fundamental importance of the parallel postulate.

5.4. Consider the two right angles case and suppose the two lines meet on one side. Argue that they should then meet on the other side also, and that this is a contradiction.

5.5. Explain how to construct a parallel to a given line through a given point based on this knowledge.

The following is an important application of the theory of parallels.

5.6. Consider an arbitrary triangle and draw the parallel to one of the sides through the third vertex. Use the resulting figure to prove that the angle sum of a triangle is two right angles.

Note that the story we have told about parallels could also have been traced backwards: starting with the angle sum theorem of a triangle, and trying to reduce it to simpler and simpler facts, we could have asked ourselves what the ultimate assumptions are that one needs to make in order to prove this theorem. This is indeed one of the great strengths of Euclid: he not only demonstrates his theorems in the sense that he convinces the reader that they are true; he also shows what, precisely, are the fundamental assumptions on which the entire logical edifice of geometry rests. The latter is no easy task; as we saw in problem 5.2 it is easy to fool yourself about what logical assumptions you are really making when you are reasoning about "obvious" things (though admittedly problem 5.1b shows that Euclid himself was not perfect in this regard either).

This kind of logical reduction of geometry to its ultimate building blocks serves no practical purpose; it is of concern only to those who take a philosophical interest in the nature and foundations of mathematical knowledge. But there were good reasons for people to see a philosophical puzzle here. What indeed is the nature of mathematics and its relation to the world? It seems that I can sit in a dark cave and prove mathematical theorems, for example that the angle sum of a triangle is two right angles, which then turn out to be true when tested on actual triangles in the real world. How can this be? In addition to this epistemological puzzle, there were paradoxes such as Zeno's rather geometrical proofs that motion is impossible and the discovery, discussed in §2, that $\sqrt{2}$ is in a sense "not a number."

Against this backdrop we can perhaps understand why the time was ripe for someone like Euclid to write a rigorous treatise systematising all of geometry. Euclid's *Elements* was not of interest for the store of theorems it proved; most of those had been known for a long time, often centuries. No, its real contribution lay in the stringent logical organisation of this material which revealed its ultimate foundations and building blocks, and thereby provided a vision of the very nature of mathematical knowledge.

The fourth postulate is the only one we have not discussed so far. It is not so central, though it is more insightful that it looks.

5.7. Argue that the fourth postulate is false on the surface of a cone. Thus it is basically a "flatness" postulate.

5.8. In the Declaration of Independence of the United States, the first part of the second sentence has a very Euclidean ring to it. It is reminiscent of one of Euclid's postulates in particular—which one? There are in fact several further allusions to Euclidean rhetoric in this document.

Figure 8: Artist's impression of Euclid with his iconic compass.

Figure 9: 13th-century illustration of God designing the world using a compass.

§ 6. Ruler and compass

As seen above, Euclid's geometry is essentially the geometry of ruler and compasses. There are many reasons why these tools form a beautiful foundation for geometry:

- Theoretical purity: line and circle, straightness and length.
- Practical simplicity: both can be generated by e.g. a piece of string.
- Practical exactness.
- Inclusive: $+$, $-$, $\times$, $\div$, $\sqrt{}$, and the regular polyhedra are all subsumed by ruler and compasses.
- Natural motion. In Aristotelian physics, earth and water want to go straight down, fire and air want to go straight up, and heavenly bodies are made of a fifth element that wants to go in circles.

6.1. In this problem we shall show that the operations $+, -, \div, \times, \sqrt{}$ can be carried out with ruler and compasses.

(a) Explain how to add and subtract, i.e., given line segments of lengths a and b, how to construct line segments of length $a+b$ and $a-b$.

Note that it is the compasses and not the ruler that enable us to treat length. Only the compass can "store" a length and transfer it to a different place.

(b) It may seem that this construction is based on nothing but some of Euclid's postulates—which ones?

Actually these axioms themselves are not enough, because Euclid's compasses "collapse" when lifted. Thus we could draw a circle with radius b centered at either of the endpoints of the line segment b, but we cannot set the compasses to this length, then lift it up and put it down at the end of a. Or rather, the axioms do not tell us that we can do this, so Euclid has to prove as a proposition that this can be done.

(c) Which of Euclid's propositions enables us to "transfer" the length b to the end of segment a? Hint: start reading Euclid's propositions from the beginning; this construction occurs very early.

Now multiplication. Given line segments $1, a, b$, to produce a line segment ab, we proceed as follows. Make 1 and a the legs of a right triangle.

(d) Explain how this is done with ruler and compasses.

Then extend the side of length 1 to a side a length b

(e) Explain how this is done with ruler and compasses.

Complete a right triangle similar to the first with this new segment as one of its legs.

(f) What is its remaining leg?

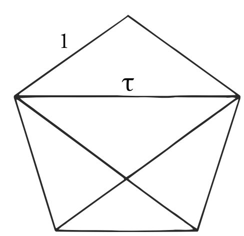
(g) Explain how to construct a/b. Hint: division is the inverse of multiplication.

To construct $\sqrt{a}$ (cf. *Elements*, II.14), draw a circle with diameter $a+1$ and draw the perpendicular at the dividing point between the a and 1 segments.

(h) Explain how $\sqrt{a}$ can be obtained from the resulting figure. Hint: recall Thales' Theorem of problem 3.9.

6.2. In this problem we shall show that the regular pentagon is constructible by ruler and compasses. Equilateral triangles, squares, and regular hexagons are easily constructed, as are any regular polygons with twice as many sides as a previously constructed polygon. Some regular polygons, such as the heptagon (7-gon), cannot be constructed with ruler and compasses.

The pentagon is an interesting case since it is needed for one of the regular polyhedra, the dodecahedron, and since its construction is non-trivial. Euclid's construction of the regular pentagon (IV.11) is much too complicated for us to go into here, but using the correspondence between algebra and geometric constructions developed above we can see that it must be possible to construct the regular pentagon.



(a) Use the symmetry of the regular pentagon to find similar triangles implying $\tau = \frac{1}{\tau - 1}$.

(b) Solve this equation for τ. Note that τ is constructible by problem 1.

(c) Give step-by-step instructions for how to construct a regular pentagon given line segments of length 1 and τ.

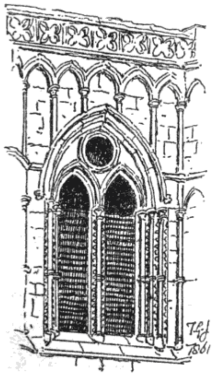


Figure 10: A typical window design in Gothic architecture.

6.3. Figure 10 shows a common design of Gothic windows.

   (a) How can you construct such a design using ruler and compasses?

   (b) Name one proposition from the *Elements* that may have suggested the basic idea of such a design. Hint: start looking at the beginning.

   The Gothic style of architecture arose in the early 12th century, within a decade or two of the first Latin translation of Euclid's *Elements*. It is not known whether, or to what extent, this was a case of cause and effect. But we do know that architects of this era showed the utmost reverence for geometry in general, as seen in the reading from Simson.

§ 7. Conic sections

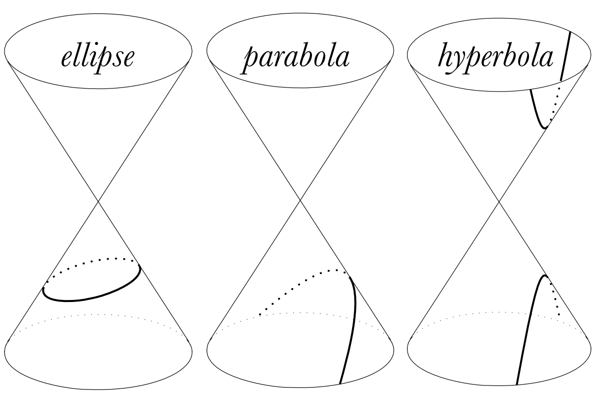
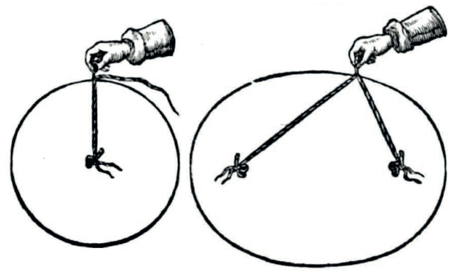


Figure 11: The conic sections.

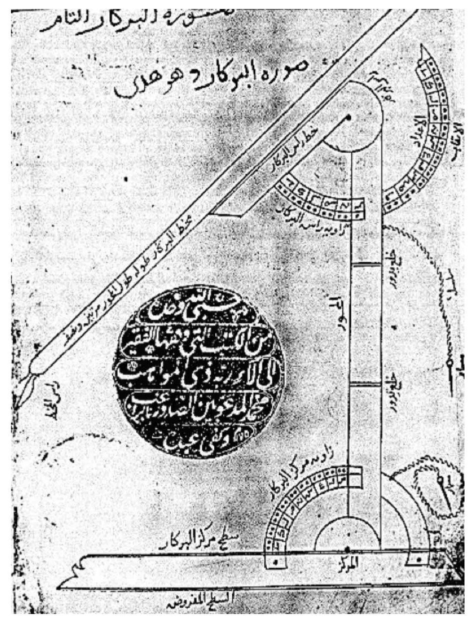
The next step beyond lines and circles are the conic sections, i.e., the curves that arise when a cone is cut by a plane. There are three fundamentally different kinds of conic sections: ellipse, parabola, and hyperbola. (Figure 11.) These terms mean roughly "too little," "just right," and "too much," respectively, which makes sense as characterisations of the inclination of the cutting plane, as we see in the figure.

7.1. Discuss how this relates to the meaning of the words ellipsis (the omission from speech or writing of words that are superfluous; also the the typographical character "…"), parable (a simple story used to illustrate a moral or spiritual lesson), hyperbole (exaggerated statements or claims not meant to be taken literally).

Many of the important reasons we used to justify ruler and compasses generalise directly to conic sections, e.g., construction by strings:



Even the compass can be generalised to draw conics, as shown in figure 12.



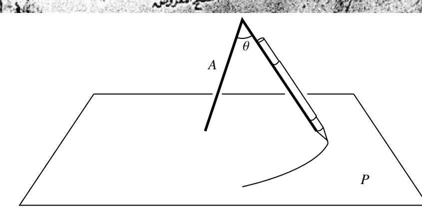


Figure 12: Generalised compass for drawing conic sections, as described by al-Kuji, c. 980, and a schematic illustration from Stillwell, *Mathematics and Its History*.

7.2. Explain how this instrument works.

7.3. This instrument is not mentioned until al-Kuji around 980, but argue that this construction is implicit already in the Greek tradition. Hint: what is the definition of a cone?

Although not known to the Greeks, the natural motion argument also generalises to conic sections, for projective motion is parabolic and planetary orbits are elliptical, as discovered by Galileo and Kepler respectively in the 17th century. Another 17th -century discovery that reveals the conics as the natural successors of lines and circles is the algebraic geometry of Descartes.

7.4. (a) In three-dimensional space, the equation $x^2 + y^2 = z^2$ represents a cone. Why?

(b) This implies that “conic sections” are curves of degree two. Consider for example the plane $y = 1$. Why is $y = 1$ a plane? What will be its intersection with the cone? What type of conic section is it?

(c) The Greeks had no algebra or coordinate systems, and yet they too knew that conics were of “degree two.” Explain how by giving a purely geometrical definition of this concept.

7.5. “Gnomon” is a fancy word for a stick standing in the ground. The tip of its shadow traces a curve as the sun moves, as indicated in figure 13.

(a) What type of curve is it?

(b) How can you find north by using the stick and the curve?

(c) At the spring and autumn equinoxes the shadow cast by the gnomon is a straight line. Explain why.

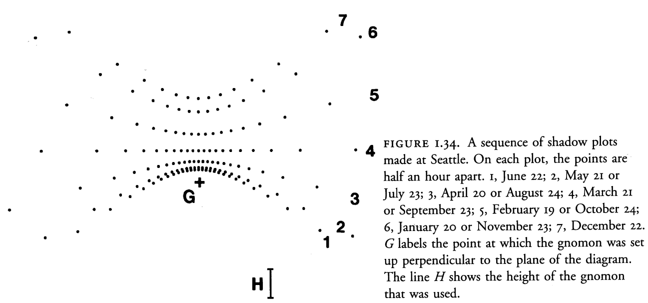


Figure 13: From Evans, *History and Practice of Ancient Astronomy*.

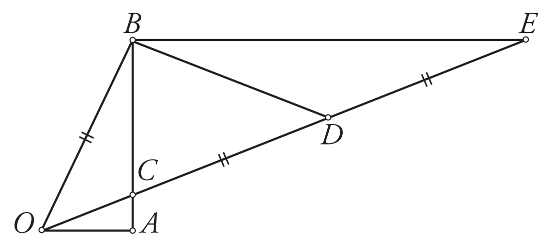
§ 8. Three classical construction problems

Three classical construction problems dominated in large part the development of Greek geometry: the duplication of the cube (finding a cube with twice the volume of a given cube), the quadrature of the circle (finding a square with area equal to a given circle), and the trisection of an angle (dividing an angle into three equal pieces). And it is with good reason that these problems were seen as fundamental. They are very pure, prototypical problems—not to say picturesque embodiments—of key concepts of geometry: proportion, area, angle. The doubling of a plane figure, the area of a rectilinear figure, and the bisection of an angle are all fundamental results that the geometer constantly relies upon, and the three classical problems are arguably nothing but the most natural way of pushing the boundaries of these core elements of geometrical knowledge. The great majority of higher curves and constructions studied by the Greeks were pursued solely or largely because one or more of the classical construction problems can be solved with their aid.

A strong case can be made that even conic sections were introduced for this reason, even though other motivations may appear more natural to us, such as astronomical gnomonics or perspective optics.

8.1. Making a cube twice as voluminous as a unit cube is obviously equivalent to constructing $\sqrt[3]{2}$. Show that this can easily be accomplished assuming that the hyperbola $xy = 2$ and the parabola $y = x^2$ can be drawn.

For trisecting an angle, one of the Greek methods went as follows.



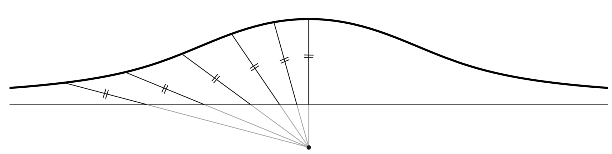
Consider a horizontal line segment OA. Raise the perpendicular above A and let B be any point on this line. We wish to trisect $\angle AOB$. Draw the horizontal through B and find (somehow!) a point E on this line such that when it is connected to O the part EC of it to the right of AB is twice the length of OB. I say that $\angle AOC = \frac{1}{3}\angle AOB$, so we have trisected the angle, as desired.

8.2. Prove that $\angle AOC = \frac{1}{3}\angle AOB$. Hint: Consider the midpoint of D of EC. It may help to draw the horizontal through D and see what you can infer from this.

But how exactly are we supposed to find the point E? This can in fact not be done by ruler and compass only.

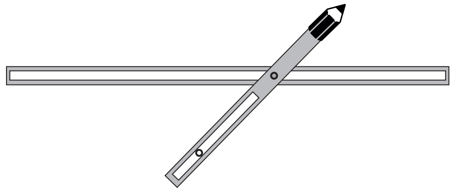
8.3. Argue, however, that it can be done if we are allowed to make marks on our ruler, and then fit the marked length into the figure by a kind of trial-and-error process. (This is called a neusis construction.)

8.4. Explain how E could also be found if we could construct curves like this:



This is called a conchoid. It was invented by Nicomedes, who also showed how it could be constructed by an instrument.

8.5. Explain how to build such an instrument. Hint:



8.6. Build such an instrument for yourself and use it to trisect an angle.

Hint: Hardware stores sometimes have tools consisting of linked rulers—sometimes called a "templater"—which are very suited for this purpose. Also, as a plane of construction it is useful to use a large sheet of very thick paper. To mark points one may use flat-headed nails piercing through the paper from below.

§ 9. Trigonometry

The history of trigonometry is the history of measuring heaven and earth. Regiomontanus called his book *De triangulis omnimodis* (1464) "the foot of the ladder to the stars."

9.1. Synopsis of Aristarchus' work *On the distances and sizes of the sun and moon* (c. -270).

Notation: E, M, S are the centers of the earth, moon and sun respectively, and E', M', S' are points on their apparent perimeters.

(a) The ratio of the distances from the earth to the moon and from the earth to the sun can be determined by measuring the angle MES at half moon. For at half moon the angle EMS=90° and the angle MES is measurable, so we know all angles of this triangle and thus the ratios of its sides.

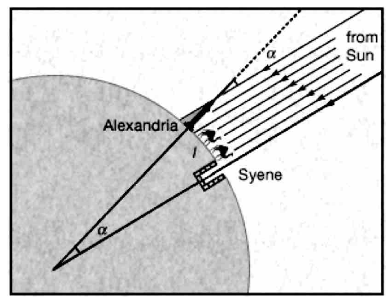
(b) The ratio of the sizes of the moon and the sun can then be inferred at a solar eclipse. For at a solar eclipse, the moon precisely covers the sun. Thus EMM' is similar to ESS', with the scaling factor discovered above, i.e. SS':MM'::ES:EM.

(c) The ratio of the distance of the moon to its size can be inferred from its angular size. For the angle EMM'=90° and the angle MEM' is measurable, so we know all angles of this triangle and thus the ratios of its sides.

(d) These distances can be related to the radius of the earth at a lunar eclipse. For the shadow that the earth casts on the moon is about two moon-diameters wide. To incorporate this information into a similar triangles setup, let O be the point beyond the moon from which the earth has the same angular size as the sun (i.e. precisely blocks out the sun). Then SS':EE'::OS:OE. Now the algebra gets a little bit involved. We want to know the LHS so we have to reduce the RHS to a number, which we will do by expressing both OS and OE in terms of OM. From above we know SS':MM', and now we have OS:OM::SS':2MM', which enables us to express OS in terms of OM. To express OE in terms of OM we first note that OE=OM+EM. From above we know how to express EM in terms of ES, or, if we prefer, MS. But again from OS:OM::SS':2MM' we know OS=OM+MS in terms of OM, so we know MS in terms of OM, so we are done. OS:OE is now some multiple of OM over some multiple of OM, i.e. a number, so we have found SS':EE', i.e. we have expressed the size of the sun (and thereby the size of the moon, of course) in terms of the size of the earth.

Aristarchus thus used the earth to measure the heavens. It remained only to determine the size of the earth itself. This was done soon thereafter by Eratosthenes.

9.2. Explain his method on the basis of the figure.



(Figure from *Encyclopaedia Britannica*.)

The contrast between the arguments of Aristarchus and Eratosthenes, though a mere generation apart, can be seen as reflecting a cultural shift in Greek antiquity. In the age of Socrates, Plato, and Aristotle, Athens was the cultural center. This was an era of abstract philosophising, of figuring things out from your armchair. We see a taste of this in the readings from Plato. Aristarchus was born on the island of Samos in classical Greece, and his measurements of the heavens embody well the power and spirit of abstract philosophising.

A new era of Hellenistic culture, however, was initiated by the conquests of Alexander the Great. His wars spread Greek culture around the Mediterranean; to Egypt among other places, where one of the cities named after him, Alexandria, was to become the new intellectual capital. Aristotle went to Macedonia to teach the young Alexander in year −343, thus marking the boundary of the two eras. Euclid wrote his *Elements* in Alexandria around year −300, synthesising great amounts of "pure" mathematics in the classical Athenian style. Later Hellenistic mathematics tends to be more "applied," broadly speaking, perhaps in part triggered by the logistic requirements of a rapidly expanding empire. Eratostenes was born in Cyrene in northern Africa—he was a "new Greek." And indeed his measurement of the size of the earth reflects the culture into which he was born. The armchair philosophers of Athens cared little about such practicalities, but when you start conquering vast lands the question soon arises: how much is there to conquer? Or: how big is the earth anyway?

9.3. China is a big country, and it has bamboos in it. This is reflected in their methods for measuring the heavens. As Chen Zi says, "16,000 li to the south at the summer solstice, and 135,000 li to the south at the winter solstice, if one sets up a post at noon it casts no shadow. This single [fact is the basis of] the numbers of the Way of Heaven." (From *The book of Chen Zi*, in the *Mathematical Classics of the Zhou Gnomon*, compiled around the first century BCE.)

Chen Zi is referring to the point S' on the earth's surface perpendicularly beneath the sun S. Now, standing somewhere else, we erect a bamboo tube BB' so that its shadow falls at our feet O.

(a) Draw a picture of this and use it to find a formula for the distance to the sun from the earth, SS', in terms of measurable quantities. (Assume that the earth is flat.)

Now, Chen Zi continues, pick up the bamboo tube you used as the post and point it towards the sun. Suppose its diameter is just big enough for you to see the whole sun through the tube (otherwise go get a longer or shorter tube).

(b) Draw a picture of this and use it to find a formula for the diameter of the sun in terms of measurable and known quantities.

(c) The value for SS' reported by Chen Zi is 80,000 li. Does this seem accurate? (Rather than trying to look up how long a li is you can answer this on the basis of the information in the quotation above.)

9.4. While the Chinese thus utilised their benefit of having vast land and bamboo sticks at their disposal, the Muslims faced other circumstances in response to which they devised other methods. Al-Biruni (*Book of the Determination of Coordinates of Localities*, c. 1025, chapter 5) discusses a method for measuring the circumference of the earth akin to that of Eratostenes, but does not find it feasible:

"Who is prepared to help me in this [project]? It requires strong command over a vast tract of land and extreme caution is needed from the dangerous treacheries of those spread over it. I once chose for this project the localities between Dahistan, in the vicinity of Jurjan and the land of the Chuzz (Turks), but the findings were not encouraging, and then the patrons who financed the project lost interest in it."

Instead: "Here is another method for the determination of the circumference of the earth. It does not require walking in deserts."

The method is this. Climb a mountain. Let M be the mountain top, E its base, and C the centre of the earth. Now look towards the horizon and let H be the point furthest away from you that you can see.

(a) Draw a picture of this and use it to find a formula for the radius of the earth in terms of measurable quantities.

Al-Biruni did indeed carry this out:

"When I happened to be living in the fort of Nandana in the land of India, I observed from an adjacent high mountain standing west from the fort, a large plain lying south of the mountain. It occurred to me that I should examine this method there. So, from the top of the mountain, I made an empirical measurement of the contact between the earth and the blue sky. I found that the line of sight had dropped below the reference line by the amount 0;34°. Then I measured the perpendicular of the mountain and found it to be 652;3,18 cubits, where the cubit is a standard of length used in that region for measuring cloth." Al-Biruni goes on to calculate the radius of the earth from this data, which comes out as 12,803,337;2,9 cubits.

These numbers are given in a mixed notation. The integer part (before the semicolon) is given in ordinary decimal notation, while the fractional part (after the semicolon) is given in sexagesimal (base 60) form. Thus, for example, $12,803,337;2,9 = 12803337;2,9 = 12803337 + \frac{2}{60} + \frac{9}{60^2}$.

(b) Check al-Biruni's calculation. Note that some discrepancy results from imperfections in the trigonometric tables available to him.

"Cubit" means "forearm," which makes sense as a unit for "measuring cloth." Although the exact value of a cubit intended by an author is often unclear, one may generally assume it to be about 44 cm.

(c) In terms of metric units, with what accuracy does al-Biruni specify the height of the mountain?

The downside of this method is of course that it requires "a high mountain close to the seashore, or close to a large level desert." Thus, coming across such a mountain is an opportunity too good to pass up even if you are in the middle of a war:

"Abu al-Tayyib Sanad bin 'Ali has narrated that he was in the company of al-Ma'mun when he made his campaign against the Byzantines, and that on his way he passed by a high mountain close to the sea. Then al-Ma'mun summoned him to his presence and ordered him to climb that mountain, and to measure at its summit the dip of the sun."

(d) However, a mountain next to a vast, completely flat land is better than a mountain next to the sea. Why? (Hint: how do you determine the height of the mountain?)

(e) Argue on the basis of figure 14 that geographical circumstances were ideal for al-Biruni's measurements.

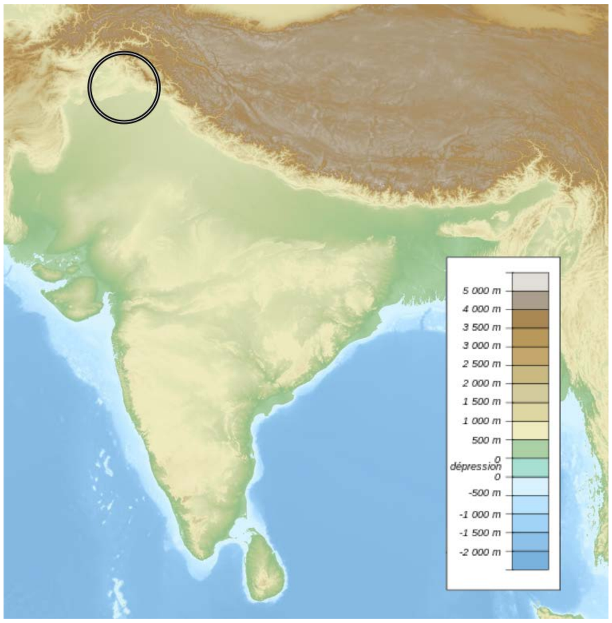
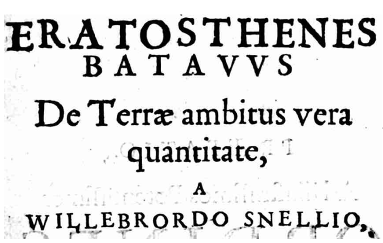


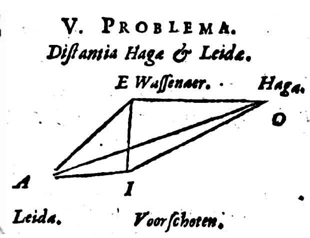
Figure 14: Location of al-Biruni's earth measurement on altitude map.

Why did the Chinese measure the heavens and the Muslims the earth? Shang Gao expresses the Chinese attitude: "one who knows Earth is wise, but one who knows Heaven is a sage." But in the Muslim world, different cultural circumstances conferred a higher status on earth-measurements:

> If the investigation of distances between towns, and the mapping of the habitable world, ... serve none of our needs except the need for correcting the direction of the qibla we should find it our duty to pay all our attention and energy for that investigation. The faith of Islam has spread over most parts of the earth, and its kingdom has extended to the farthest west; and every Muslim has to perform his prayers and to propagate the call of Islam for prayer in the direction of the qibla." (al-Biruni, *ibid.*)

In the early modern west, trigonometry was used for large-scale land surveying, which can be done by measuring one single distance between two points and then propagating this distance by triangulation by measuring nothing but angles. Snellius measured the Netherlands by this method in 1617, calling himself the "Dutch Eratosthenes" as you can see here on the title page on his book (left):





On the right is a typical figure from the book. The problem is to determine the distance from the Hague to Leiden given all the angles and one single length. Snellius measured the angles by sighting from one tall building to another, for example the Jacobstoren in the Hague, the Town Hall in Leiden, and the Nieuwe Kerk in Delft.

9.5. (a) Explain how the entire country can be measured given only angles and a single length. (Assume that the earth is flat.)

(b) This is yet another example of a trigonometric technique arising in a geographically appropriate context. How so?

The same method was employed by the French Academy in 1735 to decide between the Newtonian and Cartesian hypotheses regarding the curvature of the earth. Descartes viewed the solar system as a vortex, which led him to believe that the earth would be elongated along its axis, like a lemon. Newton argued, on the basis of his theory of gravity, that the earth was rather flattened at the poles, like an orange.

9.6. The French Academy sent expeditions to Peru and Lapland. Can you imagine why? Hint: There are at least two scientific reasons.

9.7. Carry out a triangulation project yourself. Start with a small triangle, for instance one drawn on an ordinary piece of paper. Measure only one of its sides. Then use a triangulation network to infer from this a much larger distance, such as the size of a classroom.

§ 10. Islam

During the middle ages the intellectual epicenter of the world shifted east. The Islamic world experienced a golden age around say 800–1200, while virtual barbarism reigned in mainland Europe. A number of mathematical advances in the Islamic world reflect their cultural context in an interesting way. For example, the design of mosques involved very substantial amounts of geometry for two reasons. Firstly, Muslims must pray in the direction of Mecca and mosques are aligned accordingly. Determining the direction of prayer correctly even at the outskirts of the Muslim world requires sophisticated astronomical and geographical calculations. Also, Islam prohibits the depiction of prophet Muhammed, which led to mosques being decorated with very exquisite geometrical patterns instead of figurative art.

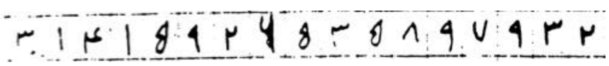
The Islamic world also played a crucial role in terms of the transmission of knowledge. A number of Greek works have come down to us only via Arabic translations and preservations. On the other side of the empire, India also provided valuable influences, such as the prototype of the "Hindu-Arabic" numeral system that we still use in the West today. Al-Biruni lived in the eastern regions of the Islamic world and knew India well. Perhaps you have heard a story involving putting grains of rice on a chessboard: one on the first square, two on the next, four on the next, and so on, doubling in each step. Al-Biruni used this example to show the superior ease with which one can calculate with the Indian numeral system as opposed to the Abjad numeral system in use at that time. Like Roman numerals, the Abjad system is not a place value system, which makes it less efficient for calculations involving large numbers. By the way, chess is originally an Indian game and rice is of course especially prominent in Indian agriculture and cuisine, so it is quite appropriate that these ingredients should figure in the promotion of Indian numerals.

10.1. Arabic is read "backwards," but numbers are written as we write them.

(a) Argue that reading a big number written in Hindu-Arabic numerals from right to left, or reading it in Roman numerals from left to right, is in a sense more natural than reading it in Hindu-Arabic numerals from left to right. Hint: what is the meaning of the first digit you encounter?

(b) Argue against this interpretation.

In 1424 al-Kashi computed π with 16-decimal accuracy using different methods. Here is al-Kashi's result in his own notation:



That's 3.1415926535897932. Note the interesting way in which our symbols for 2 and 3 are derived from their Arabic counterparts. The Arabic symbols are perhaps a more natural way of denoting "one and then some."

10.2. Thus our symbols for 2 and 3 are 90° rotations of their Arabic counterparts. This rule also works (more or less) for two other digits—which ones?

9th-century Baghdad was the centre for a massive and purposeful translation movement, aimed at translating all important works from Greece and India into Arabic. Al-Kindi was right in the thick of it, and it is surely no coincidence that he wrote a great work on cryptography in this context. For what is translation from a difficult language but a form of decryption? The *Kama Sutra* is an Indian manual for lovers. It also contains a section on encryption to facilitate secret communication between perhaps illicit lovers. The encryption is based on pairing each letter of the alphabet with another randomly selected letter. This pairing table is then used to scramble and unscramble the secret text. A similar method was used by Julius Caesar; it is easy to imagine that sensitive military and political correspondence must be kept encrypted when one is trying to run an empire by mail from some battlefront or while frolicking with Cleopatra in Egypt.

> If he had anything confidential to say, he wrote it in cipher, that is, by so changing the order of the letters of the alphabet, that not a word could be made out. If anyone wishes to decipher these, and get at their meaning, he must substitute the fourth letter of the alphabet, namely D, for A, and so with the others. (Suetonius, *Life of Julius Caesar*, 56.)

But these ciphers are not very secure at all. Al-Kindi, stimulated to think abstractly about languages and ciphers by the translation movement, described the way to crack encryptions of this kind, namely by letter-frequency analysis, and proposed better ciphers.

10.3. (a) Explain how letter-frequency analysis can be used to crack these ciphers.

(b) Consider how one might design a cipher that cannot be cracked in this way.

Linguistic influence also goes the other way. For example, the word "algebra" comes from the Arabic "al-jabr," meaning "restoration (of anything which is missing, lost, out of place, or lacking), reunion of broken parts, (hence specifically) surgical treatment of fractures" (OED).

10.4. Argue that solving quadratic equations by completing the square is indeed "al-jabr" in this sense of the word.

§ 11. Numbers in early modern Europe

We saw above that al-Biruni used a computational example involving larger and larger numbers to highlight the benefit of the Indian numeral system when he brought it to an Islamic audience. It was much the same when this system was introduced in Europe from the Arabic world. This was done by Fibonacci in his *Liber Abaci* of 1202, who introduced his famous Fibonacci numbers

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

to showcase the superior computational facility of the Hindu-Arabic numeral system.

11.1. Each Fibonacci number is the sum of the previous two. Explain how the sequence can be seen as describing a growing rabbit population, which is the example Fibonacci used.

The Hindu-Arabic numerals of course replaced the Roman numerals. In the Roman as in many other old numeral traditions, numerical values are indicated by ordinary letters. Thus 1, 2, 3, 4 was denoted by alpha (A), beta (B), gamma (Γ), delta (Δ) in classical Greek, and by alif (ا), ba (ب), jim (ج), dal (د) in the abjad system used in the Arabic world. You can still see the remnants of this in any geometry book today, for when we draw geometrical diagrams we still label our points A, B, C, etc., following the Greek tradition. But when Euclid labelled his points A, B, Γ, etc., he really meant first point, second point, third point, etc. Indeed, in the early Latin West one finds sometimes geometrical diagrams with the points labelled 1, 2, 3, etc., since they considered this the right translation of the Greek.

In any case, using the ordinary alphabet to denote numbers also has the consequence that any word is automatically associated with a number. The following example is indicative of this very widespread numerological tradition.

At the time of the Reformation in the early 16th century, Stifel proved that the reigning pope, Leo X, or Leo the tenth, was the antichrist, based on a Latin numeral interpretation of his name. "Leo X (the tenth)" in Latin capitals is LEO X DECIMVS.

11.2. Discard the letters which do not have a numerical value as Latin numerals, and write the remaining ones in descending order. What is the resulting number?

11.3. This evokes a number occurring in the New Testament, Book of Revelations, chapter 13, verse ☐, where it is called the number of the ☐.

This may strike you as a crackpot argument against the pope, and "yet this is the man who, in the next few years, produced some of the most original and vigorous mathematical works to be found in the 16th century," as noted by Smith, *History of Mathematics*, vol. 1, Dover, 1958, p. 328.

Note also how this story reflects two of the main tenets of the Reformation: distrust of the pope and the primacy of the words of the Bible. The Bible was then just recently made available to people at large due to the first translations into common languages and the invention of printing. Thus it very much fit the Zeitgeist to use the very words of the Bible itself to stick it to the authorities who used to have a monopoly on interpreting this work.

§ 12. Logarithms

Logarithms were first developed in the early 17$^{\text{th}}$ century as a means of simplifying long calculations. Long calculations were involved for example in navigation which was of increasing importance in this era. Indeed, the first ship of slaves from Africa to America set sail only four years after the publication of the first book on logarithms.

Logarithms simplify calculations by turning multiplication into addition: $\log(ab) = \log(a) + \log(b)$. This saves an incredible amount of time if you have to do calculations by hand, since it is so much easier to add than to multiply. Not long ago, before the advent of pocket calculators, people still learned logarithms for this purpose in school. Indeed, whenever you go to a used bookstore and look at the mathematics section you almost always find many tables of logarithms published some fifty or sixty years ago.

The inventor of logarithms introduced them as follows:

> There is nothing (right well beloved Students in the Mathematickes) that is so troublesome to Mathematicall practice, not that doth more molest and hinder Calculators, then the Multiplications, Divisions, square and cubical Extraction of great numbers, which besides the tedious expence of time, are for the most part subject to many slippery errors. I began therefore to consider in my mind by what certain and ready art I might remove those hindrances. And having thought upon many things to this purpose, I found at length some excellent brief rules to be treated of (perhaps) hereafter. But amongst all, none more profitable than this which together with the hard and tedious multiplications, divisions, and extractions of roots, doth also cast away from the work itself even the very numbers themselves that are to be multiplied, divided and resolved into roots, and putteth other numbers in their place which perform as much as they can do, only by addition and subtraction, division by two or division by three. (John Napier, *A Description of the Admirable Table of Logarithms*, 1616.)

This passage expresses the original purpose of the "laws of logarithms" that you have probably been taught from a very different point of view:

$$\log(xy) = \log(x) + \log(y)$$

$$\log(x/y) = \log(x) - \log(y)$$

$$\log(x^y) = y\log(x)$$

12.1. Suppose you have a table of all numbers x and their corresponding logarithms $\log(x)$. Explain how to compute xy without multiplying, x/y without dividing, and $\sqrt{x}$ without using a root extraction algorithm.

Indeed logarithms "doubled the lifetime of the astronomer," as Laplace put it. A similar endorsement is this:

> The Mathematics formerly received considerable Advantages … by the Introduction of the Indian Characters … yet hat it since reaped at least as much from the Invention of Logarithms. … By their Means it is that Numbers almost infinite, and such as are otherwise impracticable, are managed with Ease and Expedition. By their assistance the Mariner steers his Vessel, the Geometrician investigates the Nature of the higher Curves, the Astronomer determines the Places of the Stars, the Philosopher accounts for the Phenomena of Nature; and lastly, the Usurer computes the Interest of his Money. (John Keil, *A Short Treatise of the Nature and Arithmetick of Logarithms*, 1733.)

§ 13. Renaissance

Greek treatises on geometry are often hopelessly convoluted. See for example figure 15. "Intricacy and surprise govern the arrangement of the text," as the editor of the modern edition of Archimedes' works says. 17$^{\text{th}}$ -century mathematicians were convinced that these treatises were deliberately opaque and that the Greeks had a secret method of discovery which they did not reveal. In the reader I quote many expressions of this idea.

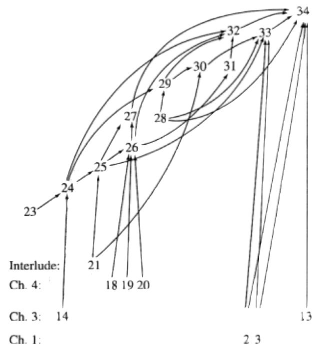


Figure 15: Dependency diagram for the very complex chain of propositions involved in Archimedes' derivation of the volume of a sphere. From Netz (ed.), *The Works of Archimedes*, Volume 1.

The suspicions of these mathematicians were dramatically vindicated when a long-lost treatise by Archimedes was discovered in 1906. In this treatise Archimedes does indeed reveal a "secret" method of discovery that has quite a bit in common with the calculus of the 17^{th} century. We shall now have a look at his proof in slightly modernised terms.

13.1. Archimedes' long lost treatise uses the "law of the lever" to arrive at geometrical results. This law states, in effect, that the lever multiplies the effect of a force by the length of the lever arm from the fulcrum to the point where the force is applied. Thus we can lift a stone with, say, a three times smaller force than that required to lift it directly by using a lever with a three times longer arm on our side than on the stone's side.

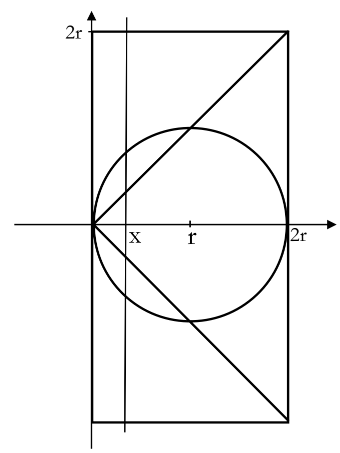
As Archimedes writes in the preface to his treatise: "Certain things first became clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards because their investigation by the said method did not furnish an actual demonstration. But it is of course easier, when we have previously acquired, by the method, some knowledge of the questions, to supply the proof than it is to find it without any previous knowledge."

In proposition 2 of his treatise, Archimedes reduced the complicated volume of a sphere to the easy volumes of a cylinder and a cone by proving that the lever arrangement below is in equilibrium: a cylinder with base radius $2r$ and height $2r$ placed with its center of mass r to the left of the center balances a cone of base radius $2r$ and height $2r$ and a sphere of radius r placed with their centers of mass $2r$ to the right of center.

(a) Express the fact that these bodies are in equilibrium as an equation. The volumes of the cylinder and cone (one third of the circumscribing cylinder) are considered known, but the volume of the sphere should be left as an unknown. (Assume, of course, that all bodies have the same uniform density.)

(b) Solve for the volume of the sphere. (This should of course give you the famous formula for the volume of a sphere.)

We shall now prove that these bodies are indeed in equilibrium. First arrange the bodies as follows: sphere with midpoint $(r, 0, 0)$; cylinder with same midpoint and its lid and bottom parallel to yz-plane; cone with bottom parallel to yz-plane and vertex at origin.



We shall cut the bodies into infinitely thin vertical slices and prove the equilibrium slicewise. Since a slice is infinitely thin, its weight is completely determined by the area of the cut, and not by the large scale shape of the object it came from.

(c) Fill in the blanks: For a given x-coordinate the cross sections (parallel to yz-plane) have the areas

$\boxed{}$ for the cylinder,

$\boxed{}$ for the cone,

$\boxed{}$ for the sphere.

(d) Now let the x-axis be a lever with the fulcrum at the origin. Prove that a slice of the cylinder, kept in its position, will balance the corresponding slices of the cone and the sphere put at $(-2r, 0)$.

(e) Conclude the proof of the volume of the sphere.

13.2. Test Archimedes's result empirically using a balance and some clay.

13.3. Archimedes would never have expressed the volume of the sphere as a "formula," but rather as a geometrical relation. Indeed, Plutarch writes of Archimedes that "although he made many excellent discoveries, he is said to have asked his kinsmen and friends to place over the grave where he should be buried a cylinder enclosing a sphere, with an inscription giving the proportion by which the containing solid exceeds the contained." What is the proportion in question?

13.4. In 1635 Cavalieri solved this kind of problem with a different method, namely the principle that if two bodies have the same cross-sectional areas at any height then they have the same volume. Using this method to find the volume of a sphere makes the above proportion especially vivid. Consider a sphere and its circumscribing cylinder. Cut two cones out of the cylinder: both with vertex at the midpoint of the cylinder, and one sharing a base with the cylinder, the other being upside-down, having the cylinder's "lid" as its base.

(a) Prove that the cross-sectional areas of these figures (at any given height) are equal.

(b) Explain how the proportion referred to by Plutarch follows from this.

13.5. Fermat's theory of maxima and minima, developed in the 1630's, is based on the idea of extrema being double roots. Say for example that we want to maximise $f(x) = x - x^2$. Fermat's method goes like this. Pick some Y smaller than the maximum. Then $Y = f(x)$ will have two solutions (one for each branch of the parabola), call them X and $X + D$. Thus $f(X) = Y = f(X + D)$.

(a) Find a simple equation relating X and D from this equality.

(b) For the maximal Y the two roots coincide (at the vertex of the parabola), i.e., the maximum corresponds to the condition $D = 0$. Use this to solve for X in your equation.

13.6. Fermat's theory of tangents was a byproduct of this method. Say for example that we want to find the tangent to $y = x^2$ at the point $(2,4)$.

(a) Draw a picture of this.

The tangent line is below the curve everywhere except at the point of tangency. In other words, among all points (x, y) on the tangent, the point of tangency minimises the quantity $x^2 - y$. Using the theory of optimisation above, we suppose, counterfactually, that there is another x-value, say $x = 2 + D$, for which the quantity $x^2 - y$ is the same as for $x = 2$.

(b) If the tangent line has y-intercept $-Y$, what is its slope?

(c) Therefore, what is its y-value when x is $2 + D$?

(d) Equate the two different expressions for $x^2 - y$ (one for $x = 2$ and one for $x = 2 + D$), and simplify.

(e) Since the minimum is actually unique, $D = 0$ after all. Plug this into your equation and solve for Y.

13.7. In 1637 Descartes published a similar method for finding normals (and thereby tangents). Suppose we seek the normal to the parabola $y = x^2$ at the point $(1,1)$. This normal is determined by its intersection with the y-axis, call it $(0, Y)$. Consider the circle centred at this point passing through $(1,1)$.

(a) Draw a picture.

(b) Write down the equation for the circle.

(c) Take its intersection with the parabola $y = x^2$ by replacing x^2 by y.

(d) Since there is only one y-value for which the circle intersects the parabola, this equation must have a double root, and thus be of the form $(y - a)^2 = 0$, i.e., $y^2 - 2ay + a^2 = 0$. What is a and a^2 in our case? Express them in terms of Y.

(e) Set $(a)^2 = a^2$ and solve for Y.

(f) Verify your answer using calculus.

13.8. When Leibniz introduced the calculus in 1684 he used it to find the tangents of curves defined as the loci of points whose distance to a fixed set of points is constant. Such curves are a kind of generalised ellipses, as indicated in figure 16. Leibniz used this example as an illustration of a problem that the calculus can handle easily but which would be almost impossible to solve by the methods of Descartes and Fermat. Explain why.

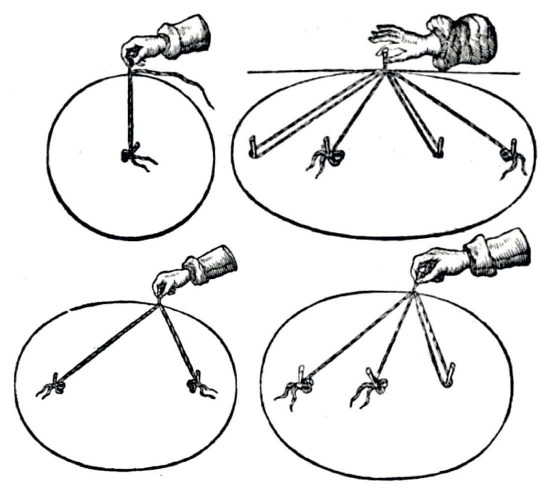
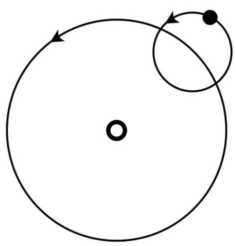


Figure 16: String constructions of circle, ellipse, and generalisations.

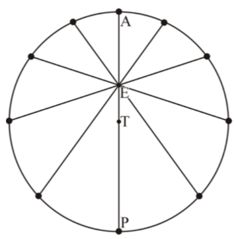
§ 14. Astronomy

As we have seen in the reader, classical cosmology is based on the mathematical-metaphysical principle that the circle is the perfect shape. The evident roundness of the sun and the moon, and the circular motions of the heavenly bodies, was seen as an expression of their divine perfection. Classical astronomy always operated within these metaphysical parameters. Though the technical task of astronomy was always to predict the position of the heavenly bodies, the theories posited for doing so had to agree with this basic cosmological framework. Ptolemy's *Almagest* (c. 150) was the definitive standard work in astronomy from Greek times to the renaissance. While Plato and Aristotle were merely philosophising conceptually, Ptolemy does the actual technical work of providing numerically explicit models for the planetary motions capable of making very accurate predictions. But nevertheless Ptolemy subscribes to the exact same metaphysical commitment, as we see in the reader.

The motions of the planets are quite complicated, however, and can by no means be seen as simple circular motions. Ptolemy therefore had to use variants on circular motion that could nevertheless be construed as being in agreement with the metaphysical principle of circular motion. Most notably he used epicycles (i.e., circles upon circles):



and equants (i.e., circular motion whose speed is uniform not as seen from the centre but as seen from some other point):



But is the use of equants not a betrayal of the principle of perfect circular motion? Copernicus thought so, and made his dislike of the equant known immediately at the very beginning of the *Commentariolus* (c. 1510):

> Ptolemy … envisioned certain equant circles, on account of which the planet never moves with uniform velocity …. A theory of this kind seemed neither perfect enough nor sufficiently in accordance with reason. Therefore I often pondered whether perhaps a more reasonable model composed of circles could be found from which every apparent irregularity would follow while everything in itself moved uniformly, just as the principle of perfect motion requires.

Copernicus is of course more famous for putting the sun at the center of the solar system. In Ptolemy's system the earth is the centre of the universe and the heavenly bodies all revolve around it. This agrees well with appearances, commonsense physics, the conception of the cosmos outlined above, and even the Bible. So it is no wonder that this was the standard view for a thousand years and more. But Copernicus turned the universe inside out. Our first instinct is perhaps to praise him for his bold and forward-thinking insight, and indeed he deserves praise. But his own motivations were for arriving at his system were quite different than we might imagine. In particular, his philosophical aversion to the equant seems to have been crucial to him. So one could make a strong case that Copernicus was not at all a revolutionary but rather an archconservative: he wanted nothing more than to go *back* to the original vision of circular motion by Plato. Thus one of the greatest scientific advances of the era was based on a reactionary philosophical principle that was universally recognised as complete nonsense just a century or so later.

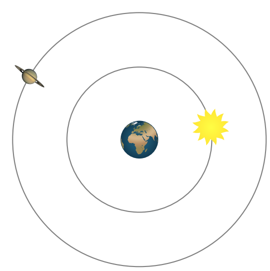
Whatever the motivations for its original discovery, the sun-centered model of the solar system has a number of advantages over the earth-centered one.

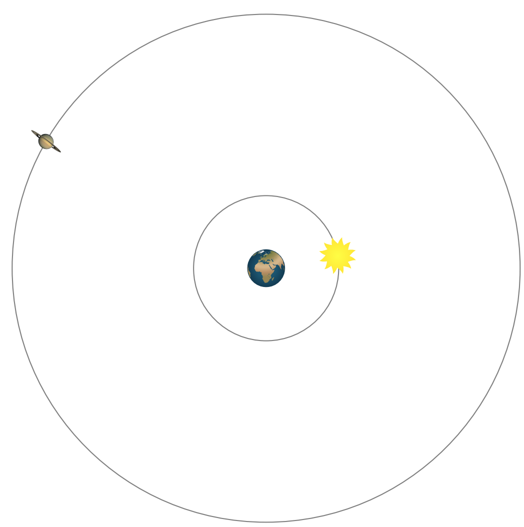
14.1. Explain why the sun-centered model gives more natural explanations of the retrograde motions of especially the outer planets.

14.2. Explain why the sun-centered model explains why the inner planets (Mercury and Venus) never deviate too far from the sun while the other planets can be found in any position relative to the sun.

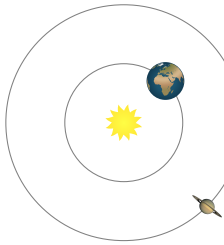
14.3. Explain why the relative distances to the planets are not determined in an earth-centered model of the solar system, but are so in a sun-centred model.

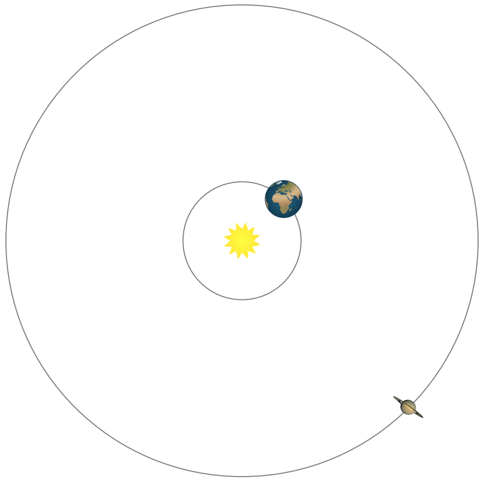
Hint: Are the following configurations observationally equivalent?





And the following two?





As Copernicus puts it, the heliocentric system "binds together so closely the order and the magnitudes of all the planets and of their spheres or orbital circles and the heavens themselves that nothing can be shifted around in any part of them without disrupting the remaining parts and the universe as a whole." For this reason he can claim triumphantly that earlier astronomers "have not been able to discover or to infer the chief point of all, i.e., the form of the world and the certain commensurability of its parts. But they are in exactly the same fix as someone taking from different places hands, feet, head, and the other limbs—shaped very beautifully but not with reference to one body and without correspondence to one another—so that such parts made up a monster rather than a man."

The Copernican system conflicts with the ancient vision of the cosmos and the polyhedral theory of the elements in a number of ways.

14.4. Outline how.

But Kepler, as we saw in §2, was fascinated by the mathematical beauty of those theories and was not about to give them up so easily. In his *Mysterium Cosmographicum* (1596), gave a very imaginative interpretation of the Copernican universe in terms of the regular polyhedra, using them to explain "the nature of the universe, God's plan for creating it, God's source for the numbers, the reason why there are six orbits, and the spaces which fall between all the spheres." See figure 17. The spheres of the planets are nested in such a way that the regular polyhedra fit precisely between them. Kepler's theory fits the facts remarkably well, and is a striking unification of the two special numbers from §2.

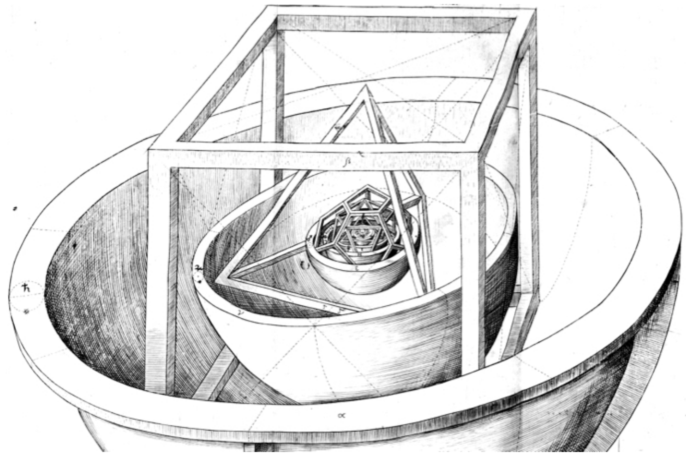


Figure 17: Kepler's theory of planetary distances.

14.5. Explain why the sun-centred model of Copernicus was crucial in two ways for Kepler to be able to posit this theory.

You might think that the "p-value" of Kepler's theory does not pass a significance test since there are so many possible permutations of the polyhedra—shouldn't one of them fit just by chance? Kepler had some dubious arguments for his particular ordering, a sample of which you will find in the readings. But even if you are not impressed by these arguments, Kepler's theory is not as arbitrary as you might think, because many permutations do not change the sizes. For example:

14.6. Show that the cube and the octahedron are interchangeable (i.e., have the same ratio of circumradius to inradius).

§ 15. Galileo

Galileo was a prominent advocate of Copernicus's heliocentric system; this even got him into hot water with the church authorities, as we learn more about in the readings. Galileo presented his case in *Dialogue Concerning the Two Chief World Systems* (1632). At the end of the work one reads:

> In the conversations of these four days we have, then, strong evidences in favor of the Copernican system, among which three have been shown to be very convincing—those taken from the stoppings and retrograde motions of the planets, and their approaches toward and recessions from the earth; second, from the revolution of the sun upon itself, and from what is to be observed in the sunspots; and third, from the ebbing and flowing of the ocean tides.

The first argument is that of problem 14.1. The third argument is that the tides can only be due to the earth's motion:

> There is nothing we can do to replicate artificially the motions of the tides, apart from moving the vessel containing the water. Surely this is enough to convince anyone that any other cause that is put

> forward to explain this effect is a vain fantasy that has nothing whatever to do with the truth?

This argument agrees very well with the mechanical philosophy and experimental spirit that we see more of in the readings. But of course it is completely wrong: the tides are caused by the gravitational attraction of the moon, and thus could just as well occur on the stationary earth of Ptolemy.

The third argument comes from telescopic observations, which Galileo was among the first to perform. One discovery was that Venus shows phases like the moon, i.e., only the half of it facing the sun is lit up.

15.1. How is this evidence for Copernicus's system? How does it fall short of a conclusive demonstration?

Thus Galileo's third argument is based on another telescopic observation, namely the sun has spots on it. As the sun rotates on its axis (with a period of less than a month), sunspots trace out latitude lines on its surface. In the course of a year, these curves are seen as alternately happy mouths, straight diagonal, sad mouths, straight diagonal, etc. This is what one would expect from a Copernican point of view if the sun's axis is inclined relative to ecliptic (i.e., the plane of the earth's orbit).

15.2. Explain.

To explain this from a geostatic point of view is more complicated since it requires the sun's axis to change inclination in a conical motion with a period of one year.

15.3. Explain.

Galileo rejects this motion of the axis as physically implausible. However, when doing so he conveniently forgets that the earth has precisely such a motion (with the same orientation; albeit a much slower one), which is the reason for the precession of the equinoxes, as Copernicus explained. Thus Galileo cannot reject as unreasonable the geostatic account of sunspot paths without simultaneously rejecting the precession of the equinoxes. Galileo was surely aware of this but suppressed it; he was an opportunist through and through.

In fact, Galileo published his sunspots argument knowing full well not only that it refutes Copernicus as well as Ptolemy but also that it was resolutely falsified by the data he himself had collected and published when he was still unaware of what his theory said he was supposed to see. Scheiner did not like the idea that the sun had blemishes on its surface, so he hypothesised that the sunspots were planets. Galileo disagreed and undertook careful observations to establish that the sunspots exhibited foreshortening effects and differences in velocity at the center as compared to the perimeter just as one would expect if the sunspots were on the surface (or atmosphere) of the spherical sun. He could then triumphantly refute Scheiner with what he called "observations and diagrams of the sunspots … drawn without a hairsbreadth of error."

Unfortunately for Galileo, he did not yet know that he was supposed to observe the sunspot paths as inclined to the ecliptic. Instead, in his resulting *Letters on Sunspots*, he asserted on the contrary that the paths of sunspots were in fact parallel to the ecliptic. When Galileo finally realised that inclined sunspot paths spoke in favour of heliocentrism, he immediately threw all his old observations "without a hairsbreadth of error" out the window and rushed the pro-Copernican argument into print. This whole business goes to show that scientific data can be a rather pliable thing, at least to an opportunist like Galileo.

Galileo also worked on mechanics. For example, have you ever considered why Newton's law $F = ma$ has acceleration in it, and not, say, velocity? In fact this has to be so because to stand still and to move with constant velocity is physically equivalent. That is, no physical experiment can tell one state from the other. This was known to Galileo, who explained it as follows in his *Dialogue*:

> Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though doubtless when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. In jumping, you will pass on the floor the same spaces as before, nor will you make larger jumps toward the stern than toward the prow even though the ship is moving quite rapidly, despite the fact that during the time that you are in the air the floor under you will be going in a direction opposite to your jump. In throwing something to your companion, you will need no more force to get it to him whether he is in the direction of the bow or the stern, with yourself situated opposite. The droplets will fall as before into the vessel beneath without dropping toward the stern, although while the drops are in the air the ship runs many spans. The fish in their water will swim toward the front of their bowl with no more effort than toward the back, and will go with equal ease to bait placed anywhere around the edges of the bowl. Finally the butterflies and flies will continue their flights indifferently toward every side, nor will it ever hap-

> pen that they are concentrated toward the stern, as if tired out from keeping up with the course of the ship, from which they will have been separated during long intervals by keeping themselves in the air. And if smoke is made by burning some incense, it will be seen going up in the form of a little cloud, remaining still and moving no more toward one side than the other.

This means that physical laws cannot speak directly about velocity. An observer on the shore thinks the guy in the ship is moving; but the guy in the ship could claim that he is in fact standing still and that it is the guy on the shore that is moving. As we just saw, no physical experiment can settle their dispute, so they must both be considered to be equally right so far as physics is concerned. Nature does not distinguish between them, so her laws must be equally true for both of them.

To illustrate this more formally, let the person on the shore be the origin of a coordinate system, and let the ship be traveling in the positive x-direction with constant velocity v. Now imagine releasing a butterfly inside the ship, in the manner described by Galileo. Suppose the butterfly moves in the x-direction only, and let $X(t)$ be its position in the coordinate system of an observer on the ship (i.e., taking a point inside the ship as the origin).

15.4. (a) Find the general formula for the position of the butterfly in the coordinate system of the observer on the shore.

(b) Express the position, velocity, and acceleration of the butterfly in terms of both coordinate systems.

(c) What is the conclusion?

It makes sense then, as you were told in physics class, that everything falls with the same gravitational acceleration, gÅ$9.8m/s^2$, at least insofar as one ignores the resistance of the air. This too is often considered a discovery of Galileo's. Indeed, in Aristotelean physics, heavier objects fall faster than light ones:

> A given weight moves a given distance in a given time; a weight which is as great and more moves the same distance in a less time, the times being in inverse proportion to the weights. For instance, if one weight is twice another, it will take half as long over a given movement. (Aristotle, *De Caelo*, I.6.)

Galileo argued against this view as follows:

> If then we take two bodies whose natural speeds are different, it is clear that on uniting the two, the more rapid one will be partly retarded by the slower, and the slower will be somewhat hastened by the swifter. … But if this is true, and if a large stone moves with a speed of, say, eight while a smaller moves with a speed of four, then when they are united, the system will move with a speed less than eight; but the two stones when tied together make a stone larger than that which before moved with a speed of eight. Hence the heavier body moves with less speed than the lighter; an effect which is contrary to your supposition. Thus you see how, from your assumption that the heavier body moves more rapidly than the lighter one, I infer that the heavier body moves more slowly. … [Since this is a contradiction] we infer therefore that large and small bodies move with the same speed provided they are of the same specific gravity [i.e., density]. (Galileo, *Dialogues Concerning Two New Sciences*, first day.)

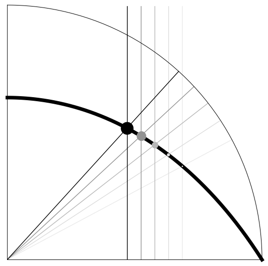
15.5. Does Galileo's argument prove that Aristotle's theory is inconsistent?

§ 16. Descartes

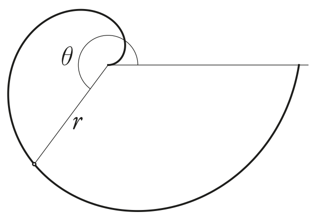
Descartes's *Géométrie* of 1637 taught the world coordinate geometry and the identification of curves with equations. However, Descartes's take on these topics is radically different from the modern view in numerous respects. In particular, Descartes did *not* argue that algebraic geometry was a replacement for classical geometry, or a radically new approach to geometry. On the contrary, he argued at great length that it was in fact *subsumed* by classical geometry, and he would never have accepted it if it wasn't. Such an attitude made perfect sense considering the unique epistemological status of classical geometry outlined in the readings.

Descartes, accordingly, began by generalising the curve-tracing procedures of Euclid and then went on to show that the curves that could be generated in this way were precisely the algebraic curves, thereby establishing a pleasing harmony between classical construction-based geometry and the new methods of analytic geometry. And with the lines, circles and conic sections of classical geometry being of degree one and two, Descartes's reconceptualisation of geometry to include algebraic equations of any degree was a natural way of subsuming and extending virtually all previous knowledge of geometry, and, at that, a way which had a definite air of seeming finality. Descartes could therefore claim with considerable credibility that only curves that could be expressed by polynomial equations were susceptible to geometrical rigour. In this way Descartes's vision of geometry masterfully combined Euclidean foundations with a bold new scope, and supplied its converts with compelling arguments as to why true geometry goes this far and no further.

Years before making his breakthroughs in analytical geometry, Descartes speculated about "new compasses, which I consider to be no less certain and geometrical than the usual compasses by which circles are traced." The key criterion for these "new compasses," according to Descartes, was that they should trace curves "from one single motion," contrary to the "imaginary" curves traced by "separate motions not subordinate to one another," such as the quadratrix and exponential curves. The quadratrix is a curve that had been considered by the Greeks; its definition uses two independently moving lines, which makes it inadmissible to Descartes:



Indeed, the ratio of the two velocities involve π, which is not known exactly, so the construction is impossible to perform in practice. The same goes for the spiral $r = \theta$.



16.1. Explain why these constructions require π to be known.

Descartes developed his single-motion criterion *before* he had the idea of a correspondence between a curve and an equation. This shows that he was very much working in the tradition of classical geometry, and that the mathematical techniques he developed were tailored to fit his philosophy of mathematics, not the other way around.

An example of his curve-tracing procedure is shown in figure 18. To find the equation of the curve traced, we can take A as the origin of a coordinate system with $AB = y$ and $BC = x$. Introduce the notation $AK = t$, $LK = c$, $AG = a$, and $m = KL/NL$. Thus t is variable while c, a, and m are constants. In terms of these quantities we can express the equations of the lines CNK and GCL, and then combine these so as to eliminate t, which gives the equation for the traced curve in terms of x, y, and constants.

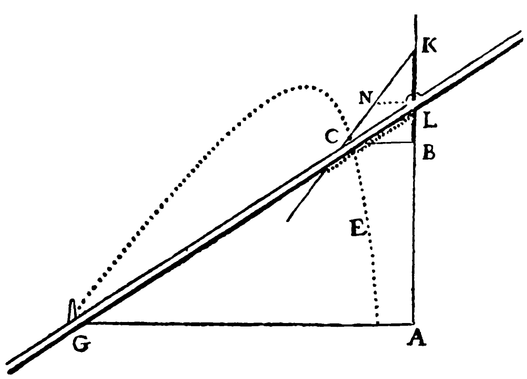


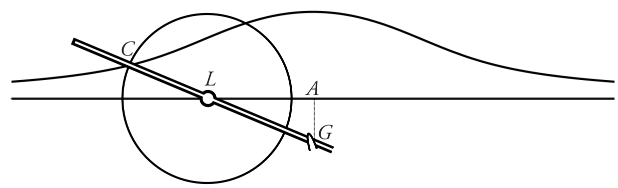
Figure 18: Descartes's method for tracing a hyperbola. The triangle KNL moves vertically along the axis $ABLK$. Attached to it at L is a ruler, which is also constrained by the peg fixed at G. Therefore the ruler makes a mostly rotational motion as the triangle moves upwards. The intersection C of the ruler and the extension of KN defines the traced curve, in this case a hyperbola.

16.2. Show how to generate the standard hyperbola $xy = 1$ using Descartes's method. That is, find a suitable choice of constants that will yield the desired curve (perhaps translated with respect to the origin, which is insignificant). Illustrate with a sketch.

16.3. Build a "new compass" for yourself and make your own $xy = 1$ hyperbola. Include the coordinate axes in your figure.

Hint: The construction tips in problem 8.6 are applicable here as well.

16.4. What curve is obtained if in Descartes's curve tracing method the line KNC is replaced by a circle with center L and radius $KL = c$?



And so it continues: once a curve has been generated this way it in turn can be taken in place of the starting curve KNC, and so on. In this way one can generate algebraic curves of higher and higher degree. Altogether, says Descartes, all algebraic curves, and nothing but algebraic curves, can be obtained in this way. This, therefore, is the domain of exact geometry according to Descartes: Euclid was right to exclude some curves (such as the spiral and the quadratrix) but wrong to limit himself to just lines and circles—geometrical rigour, according to Descartes, extends as far as all algebraic curves but no further.

§ 17. Newton's calculus

Like Descartes and Leibniz, Newton was rather ambivalent about the increasing use of algebraic methods in mathematics. As Pemberton, a contemporary of Newton, reports:

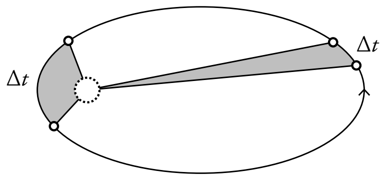
> Newton used to censure himself for not following the ancients more closely than he did; and spoke with regret of his mistake, at the beginning of his mathematical studies, in applying himself to the works of Descartes, and other algebraical writers, before he had considered the *Elements* of Euclid with that attention so excellent a writer deserves.

Indeed, in his great work *Philosophiae Naturalis Principia Mathematica* (1687), Newton shunned the calculus of formulas, even though he mastered it to perfection, and favoured instead a more geometrical style. As he says:

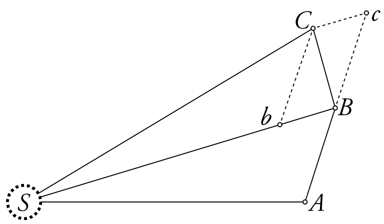
> To the mathematicians of the present century, however, versed almost wholly in algebra as they are, this synthetic style of writing is less pleasing, whether because it may seem too prolix and too akin to the method of the ancients, or because it is less revealing of the manner of discovery. And certainly I could have written analytically what I had

found out analytically with less effort than it took me to compose it.

To get a flavour of this style we may consider the very first proof in the *Principia*, namely Newton's proof of Kepler's law of equal areas. The law says that planets sweep out equal areas in equal times:



Newton's proof uses nothing but very simple infinitesimal geometry:



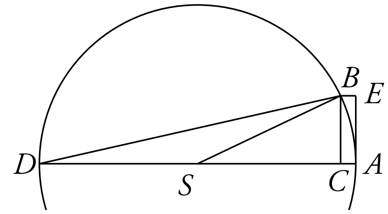
In an infinitely small period of time the planet has moved from A to B. If we let an equal amount of time pass again then the planet would continue to c if it was not for the gravity of the sun, which intervenes and deflects the planet to C. Since the time it takes for the planet to move from B to C is infinitely small, the gravitational pull has no time to change direction from its initial direction BS, thus causing cC to be parallel to BS.

17.1. Conclude the proof of the theorem.

Let us consider a second example from the *Principia*. It is the most important scientific work ever written after all. And one of its key results is the law of gravitation. The moon is kept in its orbit by the earth's gravitational pull, or so your high school textbook told you. How do you know that it is really so? How do you know that the moon is not towed about by a bunch of angels? This question doesn't seem to arise in today's authoritarian classrooms, but Newton gave an excellent answer if anyone is interested.

"That force by which the moon is held back in its orbit is that very force which we usually call 'gravity'," says Newton (Book III, Prop. IV). And his proof goes like this. Consider the hypothetical scenario that "the moon be supposed to be deprived of all motion and dropped, so as to descend towards the earth." If we knew how far the moon would fall in, say, one second, then we could compare its fall to that of an ordinary object such as an apple. Ignoring air resistance, the two should fall equally far if dropped from the same height.

Of course we cannot actually drop the moon, but with the power of infinitesimals we can deduce what would happen if we did. Here is a picture of the moon's orbit, with the earth in the center:



Suppose the moon moves from A to B along a circle with center S in an infinitely small interval of time. If there were no gravity the moon would have moved along the tangent to the circle to some point E instead of to B (BE is parallel to ASD because the time interval is infinitely small so gravity has no time to change direction).

17.2. Assuming AB being straight, use similar triangles to show that $AC/AB = AB/AD$, i.e., (diameter of the orbit)/(arc)=(arc)/(distance fallen).

Prove that ABC is similar to ABD.

17.3. Use this relation to calculate how far the moon falls in one second.

17.4. Compute how far the moon would fall if dropped at the surface of the earth, where gravity is 60^2 times stronger since the moon is 60 earth radii away.

17.5. Is the result the same as for a falling apple?

The *Principia* introduced the law of gravitation and established it as an undeniable fact. But this law is in some ways very disturbing; even Newton himself thought so (§0). This is because it seems to fly in the face of the mechanical philosophy that had gone hand in hand with all recent scientific advances. Recent science had prided itself on banishing all forms of "occult" forces in favour of concrete explanations of natural phenomena based on nothing but push-pull mechanics. But how to explain gravity in such terms? Metaphysically, it seemed as occult a force as there ever was. Newton himself and many others speculated at the time that gravity might be due to some kind of imperceptible "rain" of fine particles that drags bodies with it on their way to the center of the earth or the sun. But no coherent theory along these lines was found. In fact, an obvious aspect of the law of gravity seems to show that it is not caused by any kind of pushing on the surfaces of things; rather it somehow seems to "reach inside" objects effortlessly.

17.6. What well-known aspect of the law of gravity am I referring to?

17.7. Newton imagined that this reflected the omnipresence of God, a well-known tenet of Christianity expressed in the Bible, for instance in Jeremiah, chapter 23, verse $\boxed{}$.

§ 18. Leibniz's calculus

Leibniz was every bit as committed to the classical geometrical paradigm as Descartes was, as we see in the readings. Consequently the early calculus is permeated by geometrical constructions. As a simple illustration, consider for example how Johann Bernoulli expressed the solution to the differential equation $y' = y$.

18.1. Explain how solving $y' = y$ by separation of variables corresponds to figure 19. Areas in the same shade are equal. The point generalises to any separable differential equation.

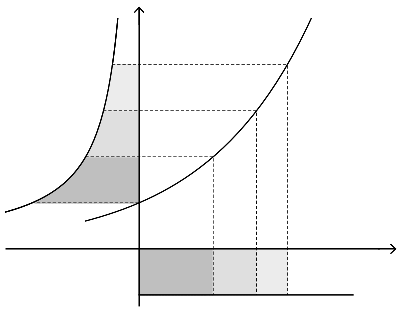


Figure 19: Geometrical interpretation of separation of variables.

So rather than looking for a "formula" for the solution, Bernoulli read the equation literally as a construction recipe. And in a sense it is not hard to understand why: what kind of "solution" is e^x anyway? It's just some arbitrary symbols. The geometrical interpretation, on the other hand, fits well with the long tradition of constructions as the bedrock of mathematics that goes back to Euclid. Leibniz too was very sensitive to this tradition, as the following example shows.

In 1691 Leibniz published a construction of the catenary (the shape of a hanging chain) corresponding to the modern formula $y = (e^x + e^{-x})/2$. His interpretation of this result differs drastically from a modern view, especially in two crucial respects:

- He never writes this formula, or indeed any formula, for the catenary. That is not what he considers a solution to a differential equation to be. Instead he "constructs" it, i.e., shows how it can be built up step by step. In this respect he is very much in line with Euclidean and Cartesian traditions, and indeed he justified his construction in such terms as we shall see.
- He sees this relationship as saying that the catenary and logarithms are essentially interchangeable. In modern terms, the function e^x is one of the most basic ingredients in the mathematician's toolbox whereas the catenary is a rather esoteric application. To Leibniz there is no such hierarchy. To him the two functions are equals. For this reason he proposes, in all seriousness, that the catenary may be used to compute logarithms: "This may be helpful since during a long journey one may lose one's table of logarithms; in case of an emergency the catenary can then serve in its place."

Leibniz's recipe for finding logarithms is shown in figure 20. Finding logarithms from a catenary may seem like an oddball application of mathematics today, but to Leibniz it was a very serious matter. Not because he thought this method so useful in practice, but because it pertained to the very question of what it means to solve a mathematical problem. Today we are so used to thinking of a formula such as $y = (e^x + e^{-x})/2$ as "the answer" to the question of the shape of the catenary, but this would have been considered a very naive view in the 17th century. The 17th-century philosopher Hobbes once quipped that the pages of the increasingly algebraical mathematics of the day looked "as if a hen had been scraping there," and what indeed is an expression such as $y = (e^x + e^{-x})/2$ but some chicken-scratches on a piece of paper? It accomplishes nothing unless e^x is known already, i.e., if e^x is more basic than the catenary itself. But is it? The fact that it is a simple "formula" of course proves nothing; we could just as well make up a symbolic notation for the catenary and then express the exponential function in terms of it. And however one thinks of the graph of e^x it can hardly be easier to draw than hanging a chain from two nails. So why not reverse the matter and let the catenary be the basic function and e^x the "application"? Modern tastes may have it that "pure" mathematics is primary, and its applications to physics secondary, but what is the justification for this dogma? Certainly none that would be very convincing to a 17th-century mind.

Thus it was with good reason that 17th-century mathematicians summarily rejected the chicken-scratch mathematics that we take for granted today. They published not formulas but the concrete, constructional meaning that underlies them. If you want mathematics to be about something then this is the only way that makes any sense. It is prima facie absurd to define mathematics as a game of formulas and at the same time naively assume a direct correspondence between its abstract gibberish and the real world, such as $y = (e^x + e^{-x})/2$ with the catenary. It makes a lot more sense to turn the tables: to define the abstract in terms of the concrete, the construct in terms of the construction, the exponential function in terms of the catenary. It was against this philosophical backdrop that Leibniz published his recipe for determining logarithms using the catenary. We see, therefore, that it was by no means a one-off quirk, but rather a natural part of a concerted effort to safeguard meaning in mathematics.

18.2. Find the value of for example log(2) by Leibniz's method using for example a neckless, a piece of cardboard, and some sewing needles.

18.3. The veracity of Leibniz's construction may be confirmed as follows. Figure 21 shows the forces acting on a segment of a catenary: the tension forces at the endpoints, which act tangentially, and the gravitational force, which is proportional to the arc s measured from the lowest point of the catenary.

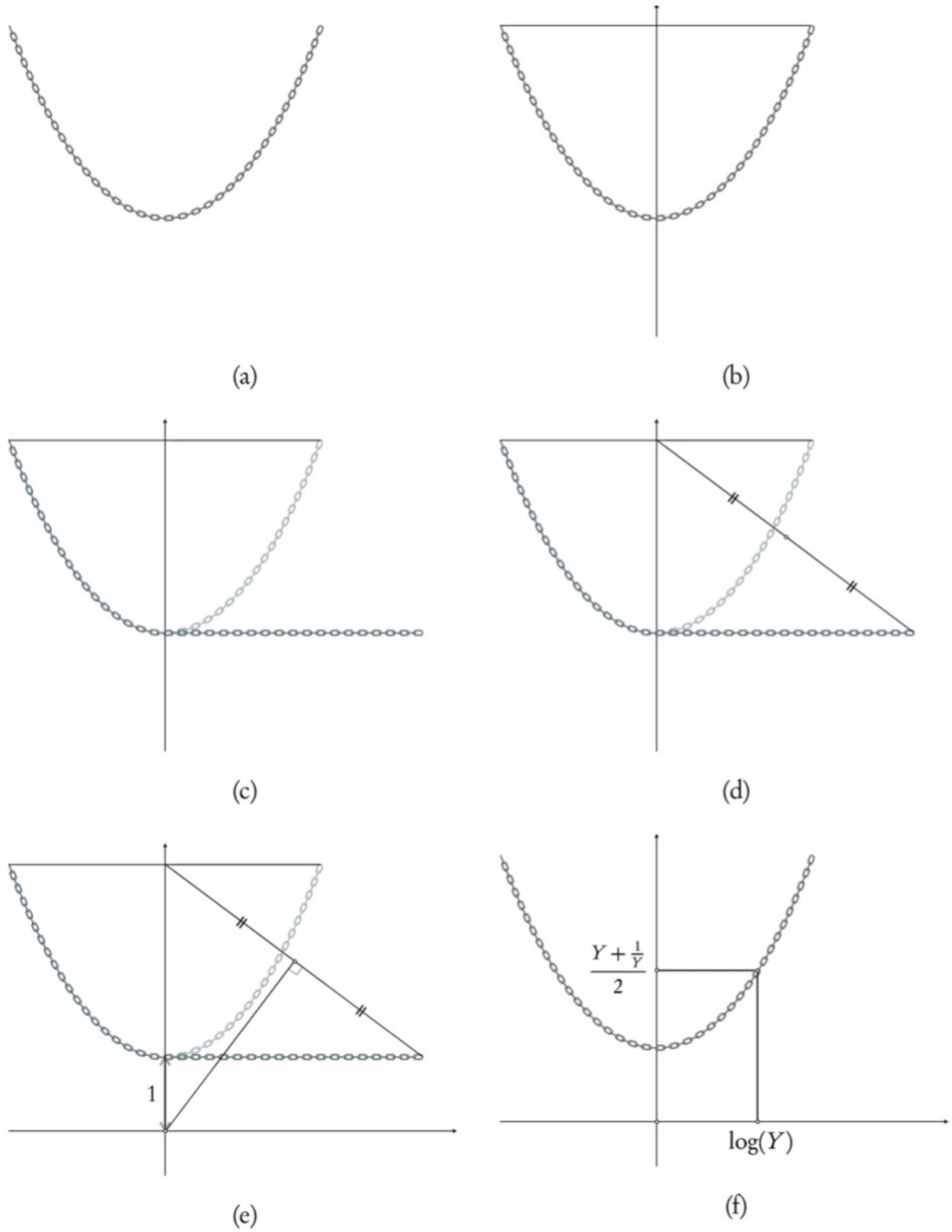


Figure 20: Leibniz's recipe for determining logarithms from the catenary. (a) Suspend a chain from two nails. (b) Draw the horizontal through the endpoints, and the vertical axis through the lowest point. (c) Pin a third nail through the lowest point and extend one half of the catenary horizontally. (d) Connect the endpoint to the midpoint of the horizontal, and bisect the line segment. (e) Drop the perpendicular through this point, and draw the horizontal axis through the point where the perpendicular intersects the vertical axis, and take the distance from the origin of the coordinate system to the lowest point of the catenary to be the unit length. The catenary now has the equation $y = (e^x + e^{-x})/2$ in the coordinate system so defined. (f) To find $\log(Y)$, find $(Y + 1/Y)/2$ on the y-axis and measure the corresponding x-value.

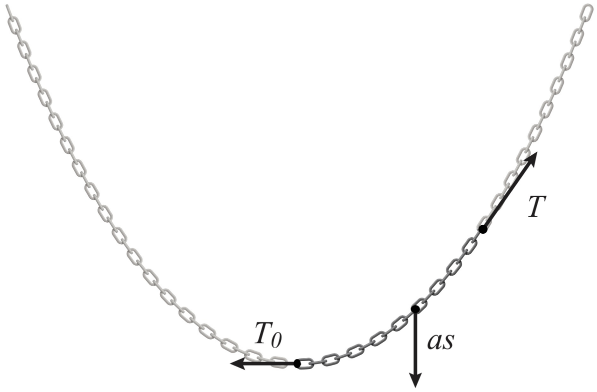


Figure 21: The forces acting on a segment of a catenary.

(a) Deduce by an equilibrium of forces argument that the differential equation for the catenary is

$$\frac{dy}{dx} = s,$$

for some appropriate choice of units.

(b) Use $dx^2 + dy^2 = ds^2$ to eliminate dx from this equation; then separate the variables and integrate. Take the constant of integration to be zero (this corresponds to a convenient choice of coordinate system).

(c) Interpret the result in terms of figure 20(e).

(d) Explain why Leibniz's construction works.

From here it is a simple matter of algebra to check the final step of figure 20, insofar as the equation $y = (e^x + e^{-x})/2$ for the catenary is known. We shall now derive this equation.

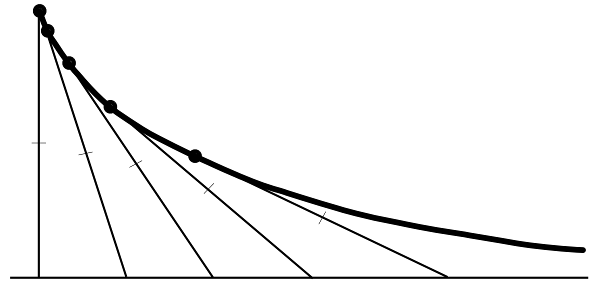
(e) In the equation you obtained in problem 18.3b, solve for s. Then substitute this expression for s into the original differential equation for the catenary.

(f) Check that $y = (e^x + e^{-x})/2$ is a solution of the resulting differential equation.

(g) Verify the final step of figure 20.

Perhaps we are too complacent today in accepting expressions like e^x or $\int dx/x$ as primitive notions just because they have a simple symbolic representation. It was different in Leibniz's day. Before long these kinds of expressions were to become accepted as legal tender but Leibniz and his contemporaries still felt obligated to fork up their actual "cash value."

18.4. Another curve related to logarithms is the tractrix, i.e., the curve traced by a weight dragged along a horizontal surface by a string whose other end moves along a straight line:



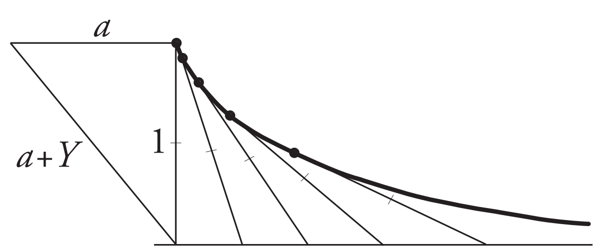
In the *physique de salon* of 17^{th} -century Paris, a pocket watch on a chain was a popular way for gentlemen to trace this curve, as shown in figure 23.

(a) Let's say that the length of the string is 1. Consider it as the hypothenuse of a triangle with its other sides parallel to the axes. Draw a figure of this triangle and write in the lengths of its sides (1 for the hypothenuse, y for the height, and the last side by the Pythagorean Theorem).

(b) Find a differential equation for the tractrix by equating two different expressions for its slope: first the usual dy/dx and then the slope expressed in terms of the triangle you just drew.

(c) Have a computer solve the differential equation for x as a function of y (using `wolframalpha.com` or *Mathematica* or similar; it is possible to solve this differential equation by hand but the calculations will be intricate). Choose the constant of integration so that the asymptote (along which the free end of the string is pulled) is the x-axis and the point $(0,1)$ corresponds to the vertical position.

The solution formula shows that the tractrix is related to logarithms. It does not reveal an easy way of finding the logarithm of some given number, but Huygens managed to turn it into such a recipe. Or rather what he did is equivalent to this. What he actually says is a bit different. He considers first this triangle, where the length of the leg a is chosen so that the hypothenuse equals this leg plus Y:



(d) Find a in terms of Y.

Next Huygens cuts off a portion of length 1 of the hypothenuse:

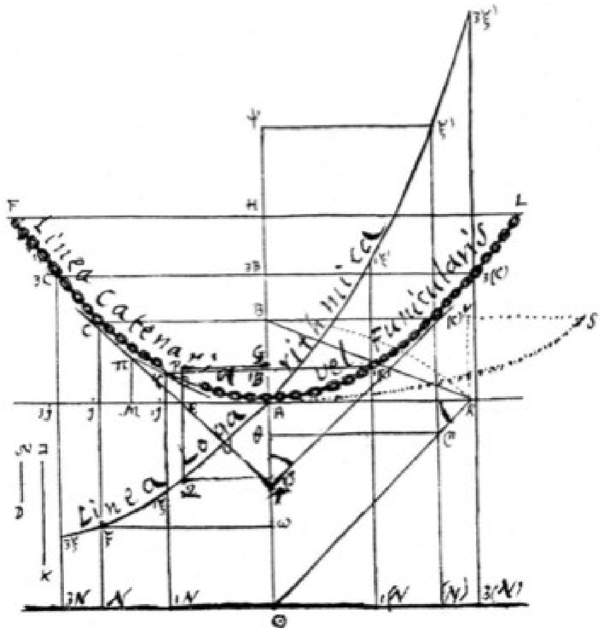
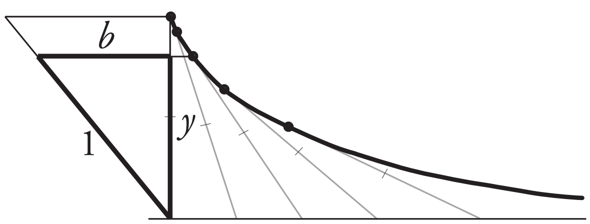


Figure 22: Leibniz's figure for his catenary construction.



(e) Find b in terms of y, and y in terms of Y.

(f) Rewrite the equation from 18.4c to obtain an expression for $\log(1/Y)$ in terms of measurable quantities (a, b, x, y).

Huygens explored the practical aspect of this construction quite thoroughly, as figure 24 shows.

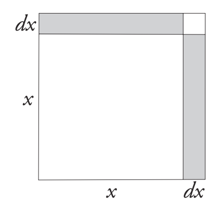
§ 19. Foundations of the calculus

The early calculus was rather freewheeling in its use of infinitesimals. The foundations for such methods eventually developed into a hot-button issue, which we shall follow in some detail in the readings. To see what all the fuss is about, we shall now have a look at the infinitesimal way of doing calculus that was the norm in the 17$^{\text{th}}$ century.

Infinitesimally speaking, to find the derivative of $y(x)$ we should:

- let x increase by an infinitesimal amount, which we shall denote dx ("d" for "difference");
- calculate the corresponding change in y, which we shall denote dy;
- divide the two to obtain the rate of change $\frac{dy}{dx}$.

In the case of $y = x^2$ this goes as follows. Suppose x increases by dx. What is the corresponding dy? It is $dy = (x+dx)^2 - x^2 = 2x\,dx + (dx)^2$ so $\frac{dy}{dx} = \frac{2x\,dx+(dx)^2}{dx} = 2x + dx$. Since dx is so small we can throw it away. Thus the derivative is $\frac{dy}{dx} = 2x$. Note that the calculations correspond to this picture:



19.1. Find the derivative of x^3 and draw the corresponding picture. Note that the derivative comes from three actual squares, a point lost on most students who learn to parrot "three x squared."

19.2. Prove the product rule $(fg)' = f'g + g'f$ in a similar way and draw the corresponding picture.

As Leibniz puts it: "It is useful to consider quantities infinitely small such that when their ratio is sought, they may not be considered zero but which are rejected as often as they occur with quantities incomparably greater. Thus if we have $x + dx$, dx is rejected. But it is different if we seek the difference between $x+dx$ and x. Similarly we cannot have xdx and $dxdx$ standing together. Hence, if we are to differentiate xy we write $(x+dx)(y+dy) - xy = xdy + ydx + dxdy$. But here $dxdy$ is to be rejected as incomparably less than $xdy + ydx$."

The integral $\int_a^b y\,dx$ means the sum (hence the $\int$, which is a kind of "s") of infinitesimal rectangles with height y and base dx:

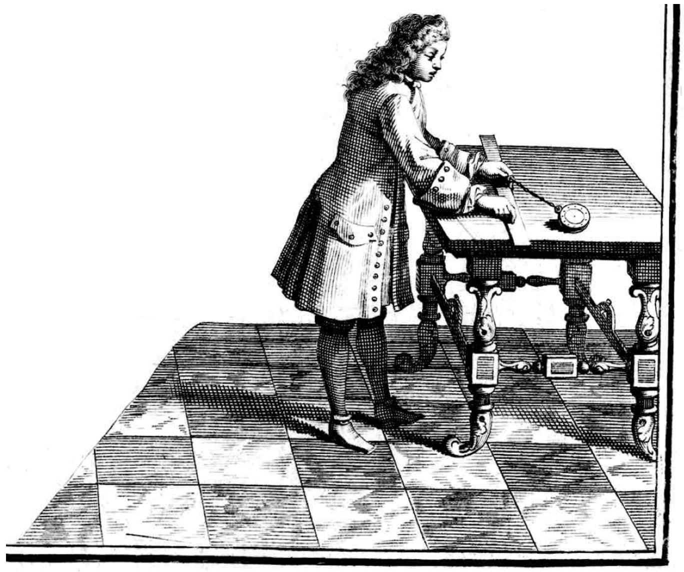
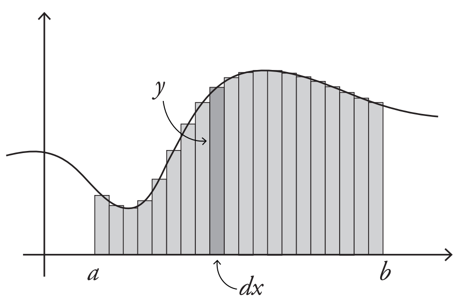


Figure 23: Tracing the tractrix by means of a pocket watch. (From Giovanni Poleni, *Epistolarum mathematicarum fasciculus*, 1729.)



19.3. Why don't the little "gaps" between the rectangles and the curve discredit the method?

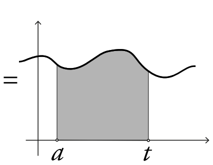
19.4. If the area under the curve is the area of the rectangles, is the length of the curve the length of the top sides of the rectangles?

The fundamental theorem of calculus says that derivatives and integrals are each other's inverses in the following ways:

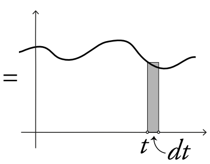
$$\frac{d}{dt}\int_a^t y(x)\,dx = y(t) \qquad \text{(FTC1)}$$

$$\int_a^b y'(x)\,dx = y(b) - y(a) \qquad \text{(FTC2)}$$

To prove FTC1 we proceed as with any derivative. In this case the variable is t and the function is $\int_a^t y(x)\,dx$.

$\int_a^t y(x)\,dx =$ area under $y(x)$ from a to $t =$ 

so if t increases by dt then $\int_a^t y(x)\,dx$ increases by

area under $y(x)$ from t to $t + dt =$  $= y(t)\,dt,$

so

$$\frac{d\int_a^t y(x)\,dx}{dt} = \frac{y(t)\,dt}{dt} = y(t),$$

which proves FTC1.

FTC2 is even easier to prove:

$$\begin{aligned}\int_a^b y'\,dx = \int_a^b \frac{dy}{dx}\,dx = \int_a^b dy \\ &= \text{sum of little changes in } y \text{ from } a \text{ to } b \\ &= \text{net change in } y \text{ from } a \text{ to } b \\ &= y(b) - y(a)\end{aligned}$$

Another way of saying this is that in order to integrate some function $f(x)$ one has only to find an antiderivative $F(x)$, i.e., a function such that $F' = f$, because then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

To Leibniz this hardly rose to the status of a theorem, let alone a "fundamental" one. He certainly never published a proof of it; in fact he barely even stated it. He was satisfied with the casual statement that "as powers and roots in ordinary arithmetic, so for us sums and differences, or $\int$ and d, are reciprocal." As far as Leibniz is concerned, the comparison is an apt one not only procedurally but also foundationally: in neither case can there be a question of proof of the reciprocal relationship; rather it is built into the very meaning of the notions involved.

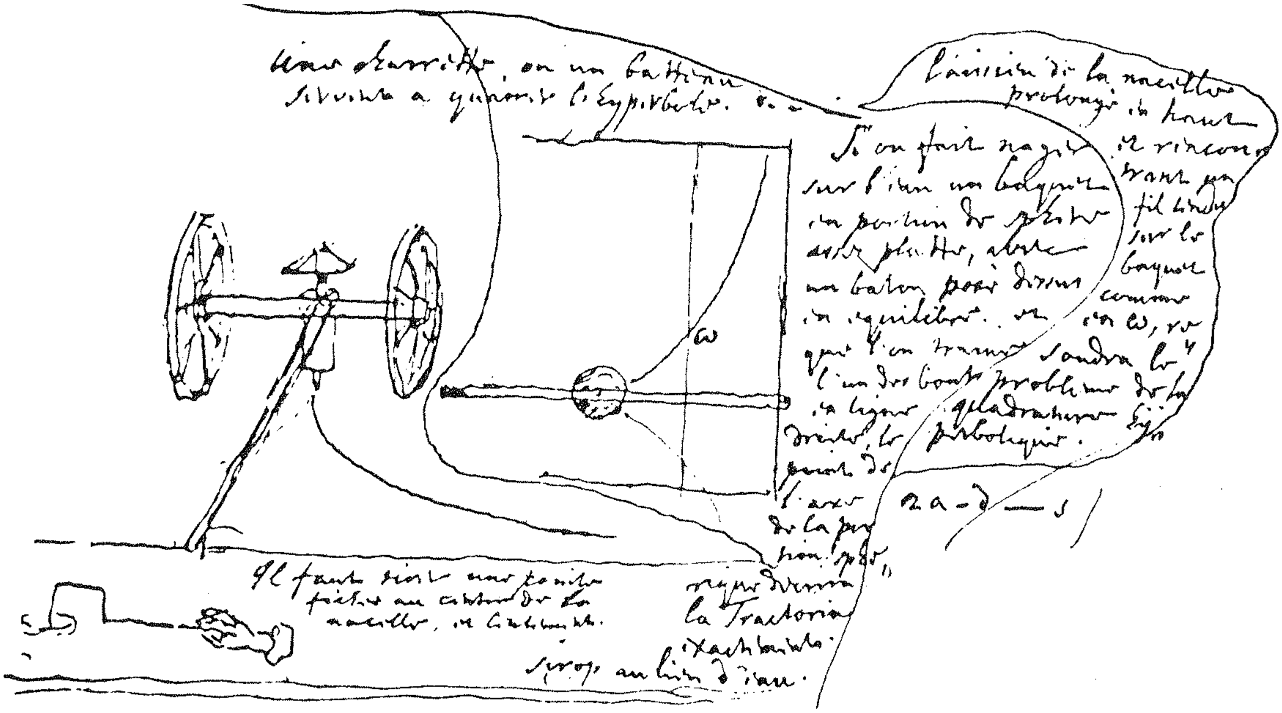
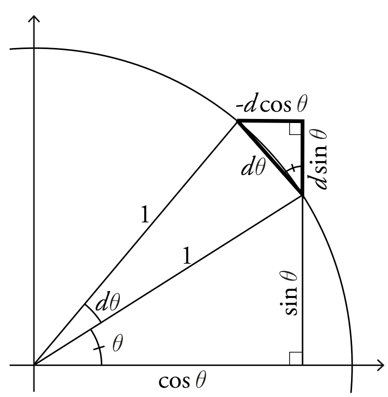


Figure 24: Detail of a 1692 manuscript by Christiaan Huygens on the tractrix. The sentence in the top left corner reads: "Une charette, ou un batteau servira a quarrer l'hyperbole" ("a little cart or boat will serve to square the hyperbola"). "Squaring a hyperbola" means finding the area under a hyperbola such as $y = 1/x$, so it is equivalent to computing logarithms, as Huygens was well aware. The bottom line reads: "sirop au lieu d'eau" ("syrup instead of water"). Syrup offers the necessary resistance and a boat leaves a clear trace in it. Using a liquid instead of a solid surface such as a table top ensures that the surface is everywhere horizontal.

19.5. (a) Explain why investigating the derivatives of sine and cosine leads to the figure below, and why the things marked as equal really are equal. Explain also why it is important that the angle is measured in radians.



(b) Find the derivatives of sine and cosine using similar triangles in this figure.

The argument in the above problem is very Leibnizian in spirit. However, explicit use and differentiation of trigonometric functions did not occur in print until a 1739 paper by Euler. These results were in effect perfectly well understood by Newton and Leibniz and others some 70 years before Euler's paper. But they did not see the need to introduce expressions like $\sin(x)$ and $\cos(x)$ into the standard arsenal of functions and study their derivatives etc. in a systematic manner. They could do pretty much everything we can do with sines and cosines, but instead of canonised notation and standard derivatives they simply expressed themselves geometrically, in terms of such-and-such an ordinate of a circle and so on. This served all their purposes perfectly well, so there was simply no need to standardise these functions. Geometrical language conveys the meaning of the results more directly; writing "$\sin(x)$" etc. would have been little more than pretentious obfuscation.

Sines and cosines solve differential equations such as the harmonic oscillator equation $\ddot{s} - s = 0$ which is so fundamental in physical theory. But this situation is simple enough that it can be described perfectly adequately in purely geometrical terms, so there was no need to write the solution as an explicit "formula." It was different for Euler. In his 1739 paper he considered a periodically forced harmonic oscillator, which we would express by the equation $\ddot{s} - s = \sin(t)$. At this point geometrical language is no longer suited for expressing the complicated solutions that arise. Euler says precisely this in a letter to Johann Bernoulli: "there appear … motions so diverse and astonishing that one is unable altogether to foresee until the calculation is finished." Only in this context did it become necessary to introduce $\sin(x)$ and $\cos(x)$ formally as functions with explicit differentiation rules and so on.

The shift from the geometrical to the analytical way of treating these kinds of situations was, then, anything but a revolution. The old, geometrical point of view held its own for a long time and was abandoned on pragmatic rather than principled grounds. It was not abandoned because it was conceptually limiting, or incompatible in principle with new directions of research, or because the new paradigm enabled some radical change of course. It was abandoned, instead, for having grown a bit too cumbersome, much like the geometrical paradigm of the Greeks eventually grew too cumbersome in practice and had to be replaced by analytic geometry.

§ 20. Power series

In modern textbooks, the theory of power series begins with a derivation of the general Taylor series, and then the series for the various standard functions are derived as special cases of this general theorem. History brings another perspective. Taylor did not publish the general Taylor series until 1715, almost half a century after Newton knew how to find the series of virtually any function. Newton and Leibniz and their followers simply had no use for the general Taylor series; they had more concrete methods which were just as powerful. In this section we shall see some of them. Furthermore we shall see Taylor's own derivation of his series, which is certainly very different from anything found in modern textbooks.

20.1. The binomial series

$$(1+x)^q = 1 + qx + \frac{q(q-1)}{2!}x^2 + \frac{q(q-1)(q-2)}{3!}x^3 + \cdots$$

was one of Newton's earliest discoveries. But to Newton the binomial series and its little brother the geometric series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

are not theorems to be proved but rather nothing but shorthand summaries of algebraic operations. Here are Newton's own words:

> Fractions are reduced to infinite series by division; and radical quantities by extraction of the roots, by carrying out those operations in the symbols just as they are commonly carried out in decimal numbers. These are the foundations of these reductions: but extractions of roots are shortened by this theorem [the binomial theorem].

Indeed, Mercator famously found a power series for the logarithm by precisely such a route, namely as follows. First note that

$$\log(1+x) = \int_1^{x+1} \frac{1}{t}\,dt = \int_0^x \frac{1}{1+u}\,du.$$

(a) Find a series for $\frac{1}{1-x}$ using long division, i.e., the kind of algorithm indicated here for 2675/25=107:

$$\begin{array}{rr|rrrr} & & & 1 & 0 & 7 \\ \cline{3-6} 2 & 5 & 2 & 6 & 7 & 5 \\ & & 2 & 5 & & \\ \cline{3-6} & & & 1 & 7 & 5 \\ & & & 1 & 7 & 5 \\ \cline{4-6} & & & & & 0 \end{array}$$

Assume that $0 < x < 1$.

(b) Plug in $x = -u$ and integrate term by term to find the series for $\log(1+x)$ in this way.

20.2. Let us use this result to illustrate another of Newton's forgotten power series techniques, namely his method of series inversion. Given a power series for y as a function of x,

$$y = ax + bx^2 + cx^3 + \ldots,$$

we compute the power series for x as a function of y, i.e., the inverse function, as follows.

- First we find the linear term. Set $x = Ay$. Substitute this into the series for y and throw away all non-linear terms. This leaves $y = a(Ay)$. Solve for A.
- Next we find the quadratic term. Set $x = Ay + By^2$ (A now being known). Substitute this into the series for y and throw away all non-quadratic terms. Solve for B.
- And so on.

(a) Find the series for $e^x - 1$ by inverting the series for $\log(1+x)$.

The series for sin, cos, and tan can all be found in the same way since their inverse functions are expressible as integrals of functions that are easily expanded as a binomial or geometric series.

20.3. Here is a way of convincing you that any function can be expressed as a power series

$$f(x) = A + Bx + Cx^2 + Dx^3 + \cdots$$

(a) Argue visually that by choosing the coefficients you can make a parabola of the form $y = ax^2 + bx + c = A(x-B)^2 + C$ go through essentially any three points but not any four.

This is because the parabola has three "degrees of freedom," i.e., you have three choices to make when picking the coefficients. Thus you can make it do three things.

(b) Adapt this argument for functions of the form $y = c$ and $y = bx + c$.

(c) Conclude that it makes sense that any function can be represented by an "infinite polynomial."

20.4. Indeed, Newton constructed such a polynomial, namely a polynomial $p(x)$ which takes the same values as a given function $y(x)$ at the x-values $0, b, 2b, 3b, \ldots$. Here is the construction. First, our polynomial $p(x)$ is supposed to have the same value as the given function $y(x)$ when $x = 0$. Therefore we should start by setting $p(x) = y(0)$. Next, we want $p(x)$ to take the same value as $y(x)$ when $x = b$. This is easily done by setting

$$p(x) = y(0) + \frac{x}{b}\big(y(b) - y(0)\big).$$

This polynomial obviously agrees with $y(x)$ when x is 0 or b. Now we need to add a quadratic term to make it agree when x is $2b$ as well. We want the new term to contain the factor $(x)(x-b)$ because then it will vanish when x is 0 or b, so our previous work will be preserved. If we set $x = 2b$ in the piece of $p(x)$ that we have so far we get

$$p(2b) = y(0) + 2y(b) - 2y(0) = 2y(b) - y(0).$$

So we want the quadratic term to have the value $y(2b) - 2y(b) + y(0)$ at $x = 2b$.

(a) Use this reasoning to write down a second-degree polynomial $p(x)$ that agrees with $y(x)$ when x is 0, b or $2b$. (Keep the factor $(x)(x-b)$ as it is, i.e., do *not* reduce the expression to the form $p(x) = A + Bx + Cx^2$.)

In the same manner we could add a cubic term to make $p(x)$ agree with $y(x)$ at $x = 3b$, and so on.

The formula becomes more transparent if we introduce the notation $\Delta y(x)$ for the "forward difference" $y(x+b) - y(x)$, and $\Delta^2 y(x)$ for the forward difference of forward differences $\Delta y(x+b) - \Delta y(x)$, etc., so that

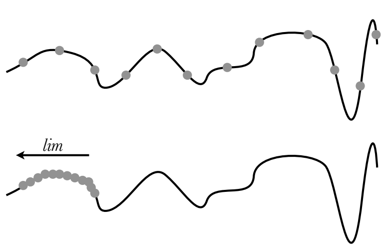
$$\begin{aligned}
\Delta y(0) &= y(b) - y(0)\\
\Delta^2 y(0) &= \Delta y(b) - \Delta y(0) = y(2b) - 2y(b) + y(0)\\
\Delta^3 y(0) &= \Delta^2 y(b) - \Delta^2 y(0) = y(3b) - 3y(2b) + 3y(b) - y(0)\\
&\vdots
\end{aligned}$$

(b) Rewrite your formula for $p(x)$ using this notation, and then extend it to the third power and beyond "at pleasure by observing the analogy of the series," as Newton puts it.

(c) Show that Taylor's series

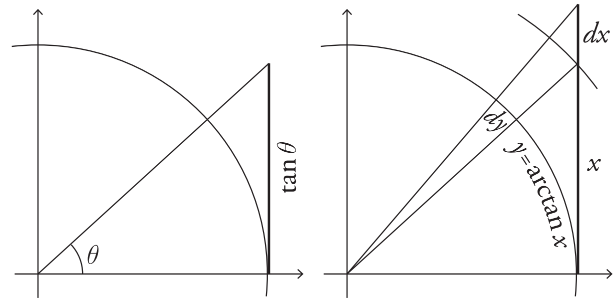
$$y(x) = y(0) + y'(0)x + \frac{y''(x)}{2!}x^2 + \frac{y'''(x)}{3!}x^3 + \cdots$$

is the limiting case of Newton's forward-difference formula as b goes to 0.



This is indeed how Taylor himself proved his theorem in 1715. The nowadays more common method of finding the series by repeated differentiation was used by Maclaurin in 1742.

20.5. On the left here is the geometrical definition of the tangent function (hence its name):



In this problem we shall investigate the inverse of the tangent function, i.e., the arctangent. For the inverse of the tangent, $\tan\theta$ is the input and θ is the output; to emphasise this we call them x and y respectively, as shown on the right. Note that since we are using radian angle measure the angle is the same thing as the corresponding arc (of course the circle is a unit circle).

Let us find the derivative of the arctangent. In other words we are looking for dy/dx. In the figure I made x increase by an infinitesimal amount dx and marked the corresponding change in y. We need to find how the two are related. To do this I drew a second circle, concentric with the first but larger, which cuts off an infinitesimal triangle with dx as its hypothenuse.

(a) Show that this infinitesimal triangle is similar to the large one that has x as one of its sides.

(b) By what factor is the second circle larger than the first? (Hint: Find the hypothenuse of the triangle with x in it. Remember that the first circle was a unit circle.)

(c) Use this to express the short leg of the infinitesimal triangle as a multiple of dy.

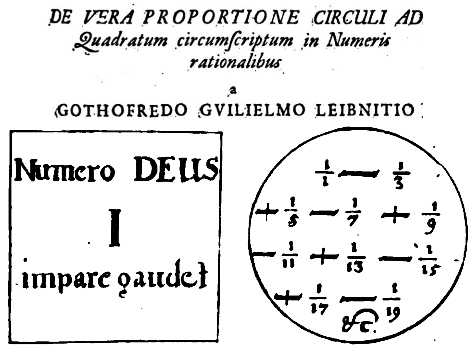
(d) Find dy/dx by similar triangles. (Check that you obtain the known derivative of the arctangent.)

(e) By the fundamental theorem of calculus, the arctangent is the integral of its derivative. Use this to find a power series for the arctangent. (To make sure that you take the constant of integration into account, check that your constant term is correct using the geometrical definition of the arctangent.)

(f) Find the value of $\arctan(1)$ in two ways: by the geometrical definition, and from the power series.

(g) Equate these two expressions for $\arctan(1)$ to find an infinite series representation for π.

When Leibniz found this series he concluded that "God loves the odd integers," as you can see in the figure below (taken from his 1682 paper).



(h) What does Leibniz's series have to do with a square of area 1, which is what Leibniz has drawn on the left?

Leibniz's series is beautiful but it is not really very efficient for computing π. Already in 1424 al-Kashi had computed π with 16-decimal accuracy using different methods.

(i) Estimate how many terms of Leibniz's series must be added together to achieve the accuracy al-Kashi had obtained already in 1424 (see §10).

§ 21. Complex numbers

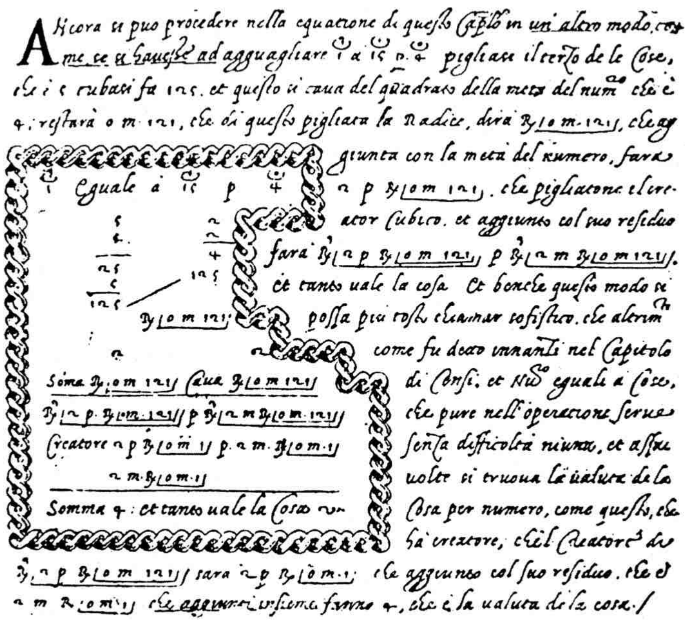


Figure 25: A page from Bombelli's *Algebra* (1572).

With complex numbers we can solve any quadratic equation, or so the textbooks tell us. But what kind of "solutions" are these weird things with i's in them anyway? Indeed, the first person to publish on complex numbers, Cardano in his 1545 treatise *Ars magna*, called them "as subtle as they are useless." This was indeed in the context of a quadratic equation. Since some students may share Cardano's lack of enthusiasm about complex numbers it may be interesting to see what compelled mathematicians to recognise the value of complex numbers despite this natural reluctance.

Bombelli was more positive towards complex numbers in 1572. But what convinced him was not the quadratic equations found in textbooks today but rather cubic ones, i.e., equations of degree 3. For cubic equations there is a formula analogous to the common quadratic formula, namely the solution of $y^3 = py + q$ is

$$y = \sqrt[3]{\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3}} + \sqrt[3]{\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3}}.$$

21.1. (a) Apply the formula to $x^3 = 15x + 4$. Before simplifying, locate (the equivalent of) the expression you obtain in Bombelli's own notation in figure 25. Hint: Bombelli denotes certain algebraic operations by the initial letters of the corresponding words.

(b) The two cube roots that arise are in fact equal to $2+i$ and $2-i$. Check this.

(c) So what solution does the formula give? Is it correct?

The conclusion is that even if you think answers with i's in them are hocus-pocus you still have to admit that complex numbers are useful for answering questions about ordinary real numbers as well.

More generally, complex numbers broke down the resistance towards them by being unreasonably effective for solving real problems. In case after case, doing algebra with complex numbers as if they were real simply works, and gives incredibly simple solutions to otherwise intractable problems. Here is an example.

21.2. Fermat claimed without proof that $y^3 = x^2 + 2$ has only one solution in positive integers.

(a) Find it.

To prove that there are no other solutions, Euler (1770) factored $x^2 + 2$ into the complex factors $(x + \sqrt{-2})(x - \sqrt{-2})$. Next he simply assumed—without worrying too much about it—that numbers of the form $n + \sqrt{-2}m$ are analogous to ordinary integers. In particular, since the left hand side is a cube, so must $(x + \sqrt{-2})(x - \sqrt{-2})$ be. And since no common factor of of these terms leaps to the eye it seems reasonable to assume that these factors are relatively prime, which means that both of them must be cubes in turn. Thus for example $x + \sqrt{-2} = (a + \sqrt{-2}b)^3$.

(b) Expand the cube and determine the possible integer values of a and b.

(c) Conclude Euler's proof of Fermat's claim.

(d) Similarly, show that $y^2 = x^5 + 5$ has a positive solution but that Euler's method fails to find it.

Another "blind faith" use of complex numbers was the following.

21.3. (a) Integrate $\arctan(x) = \int_0^x \frac{dx}{1+x^2}$ using (complex) partial fractions to obtain $\frac{i}{2}\log\frac{x+i}{x-i}$.

Johann Bernoulli did this in 1702. Basically he had no idea what a complex function was or what the logarithm of a complex number is even supposed to mean. He simply trusted the algebra and assumed that everything works the same way as for real numbers.

No one would have been very exited if this was nothing but algebraic gymnastics leading to formulas that no one knew what they meant. But Bernoulli soon figured out how to put his imaginary formula to "real" use, namely for finding multiple-angle formulas for $\tan\theta$. Let $y = \tan n\theta$ and $x = \tan\theta$. Then $\arctan y = n\theta = n \arctan x$.

(b) Use this to find an algebraic relationship between y and x.

Bernoulli admits that this formula contains "quantitates imaginarias … quae per se sunt impossibilia"—imaginary quantities which are by themselves impossible. But this, he says, is not a problem since they "in casu quolibet particulari evanescunt"—vanish in any particular case.

(c) Let $n = 3$ and find a formula for $\tan 3\theta$ in terms of $\tan\theta$ involving no imaginary quantities.

Bernoulli is quite proud to have carried out the derivation "sine serierum auxilio"—without the help of series. One benefit of this approach, he notes, is that it shows that the relationship is "semper algebraicum"—always algebraic—which is not clear from a series approach. Apparently, he considered working with "impossible quantities" a small price to pay for this added insight and simplicity.

Laplace (1810) also used complex substitutions to evaluate real integrals and called it "un moyen fécond de découvertes"—a fruitful method of discovery. But by way of justifying these methods he offered little but a vague appeal to "la généralité d'analyse"—the generality of analysis. Although his results were all correct, Poisson still found it worthwhile to rederive them by other methods, since, as he said, Laplace's reasoning was "une sort d'induction fondée sur le passage des quantités réelles aux imaginaires"—a sort of induction based on the passage from real to imaginary quantities. In a reply, Laplace agreed that the use of complex variables constituted "une analogie singulière"—a singular analogy—which "laissent toujours à désirer des démonstrations directes"—still left a desire for direct demonstrations—and he proceeded to offer some such demonstrations himself. As a result of this debate, Laplace assigned his young and ambitious protégé, Cauchy, the task of investigating the foundations of complex methods in integration. The day after his 25th birthday, Cauchy presented his "Mémoire sur les intégrales définies," aiming to "établir le passage du réel à l'imaginaire sur une analyse directe et rigoureuse"—base the passage from the real to the imaginary on a direct and rigorous analysis.

Cauchy went on to create the field of complex analysis, which blossomed into a key area of mathematics in the 19$^{\text{th}}$ century. In retrospect it is hard to imagine that such a wonderful field of mathematics was initially developed for such an esoteric purpose as to address some nagging little technical matters concerning a particular technique for dealing with certain obscure integrals which could already be dealt with by other means. But in fact it is not unusual for mathematical theories to enter the world in this backward manner, in response to some minuscule technical problem.

§ 22. Analysis in place of geometry

Blind faith in the manipulation of formulas was a successful research strategy in the 18$^{\text{th}}$ century. We already saw this with respect to complex numbers, but the point generalises. In the 17$^{\text{th}}$ -century analytical methods were primarily conceived of as a way of shortening and automatising *already existing* geometrical reasoning. An analytical proof was seen as different in form but not in principle from a geometrical one; in principle the two were intertranslatable. Descartes's attitude is typical:

> This does not make [my solution of the Pappus problem] at all different from those of the ancients, except for the fact that in this way I can often fit in one line that of which they filled several pages.

But quite soon analytical methods were found to take on a life of their own. Analytical methods began generating reasonings that had no geometrical counterpart. And, somewhat miraculously, these reasonings proved to be very reliable. Analytical methods had originally relied on their intertranslatability with geometry as the source of their credibility, and there seemed to be no reason to believe that they would not always need this crutch. But it soon became undeniable that they could stand on their own legs and even cover vast areas with ease that geometry could hardly wade through with the greatest effort.

As an illustration of the analytical magic so characteristic of the 18$^{\text{th}}$ century, consider the following early triumph of Euler (1735), which set the tone for a life's work based on bold faith in analytical methods.

22.1. (a) By considering the roots of $\sin(x)/x$, argue that its power series

$$\sin(x)/x = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots$$

can be factored as

$$\left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right)\left(1 - \frac{x^2}{9\pi^2}\right)\cdots$$

by analogy with the way one factors ordinary polynomials, such as $x^2 - x - 2 = (x+1)(x-2)$.

(b) What is the coefficient of x^2 when the product is expanded?

(c) Equate this with the coefficient of x^2 in the ordinary power series and use the result to find a formula for the sum of the reciprocals of the squares, $\sum 1/n^2$.

These kinds of examples cannot be seen as codified and streamlined geometry; they are simply inherently analytical in their very essence. The striking triumphs of these methods, therefore, force upon us the conclusion that there is something more to mathematics than the geometrical paradigm can encompass. Two possible attitudes toward this new state of affairs suggest themselves. Either we fall into an identity crisis since mathematical meaning and rigour had always been firmly anchored in the Euclidean tradition, and now these new methods are proving of undeniable effectiveness despite their dubious meaning and ontological status by any traditional standard. This reaction would certainly make sense for a philosopher concerned with the epistemology of mathematics. Alternatively, one can take a more pragmatic attitude and say that the Euclidean paradigm was justified through its triumphs in the first place, and now analytical methods have won that same warrant, so we simply admit them as equals without worrying any more about it. In other words, our attempts to mimic the Euclidean paradigm in modern times was not due to any deep-seated philosophical conviction, but was just an opportunistic attempt at mining more truths from a fruitful vein; whence it stands to reason that, at the moment it proved depleted, we did not hesitate to abandon it unceremoniously. The pragmatic attitude could not have been stomached by Descartes or Leibniz, but the new generation of mathematicians counted no philosophers among them and no qualms about taking the pragmatic route.

Lagrange was the most brazen propagandist for a wholesale break with geometry and unquestioning acceptance of analytical formulae as the new de facto subject matter of mathematics. In his *Mécanique Analytique* (1788), he put it succinctly:

> No figures will be found in this work. The methods I present require neither constructions nor geometrical or mechanical arguments, but solely algebraic operations subject to a regular and uniform procedure.

This is the direct antithesis of the view of Leibniz, you will recall. Lagrange lamented that "those who rightly admire the evidence and rigour of ancient demonstrations regret that these advantages are not found in the principles of these new methods [of infinitesimals]," and this has been the mainstream view ever since. But advocates of this view often fail to realise that it is based on a radical reconception of what "the evidence and rigour of ancient demonstrations" really consist in. Leibniz et al. were passionately dedicated to preserving "the evidence and rigour of ancient demonstrations" and stubbornly refused to budge an inch on the matter. But to them this evidence and rigour consisted first and foremost in the constructive element of the method. This is what they fought tooth and nail to preserve, and this is what Lagrange fervently purged from mathematics like so much superstition and dead weight. Indeed, if Leibniz had lived for a hundred years more one could easily imagine him criticising Lagrange's approach to the calculus in the exact same words that we saw Lagrange direct against Leibniz above. To Leibniz, it is Lagrange who has sold the soul of geometry by giving up constructions. What is at stake here is not who is rigorous, but what rigour means.

On the very same page as the above quotation Lagrange goes on to give his own supposedly more rigorous account of the application of calculus to geometry, which starts: "To consider the question in a general manner, let $y = f(x)$ be the equation of any given curve ..." In other words, the identity of curves with analytic expressions is taken for granted at the outset. This entire way of framing the question is profoundly incompatible with the 17^{th} -century interpretation of geometrical evidence and rigour. Gone is the notion that geometry constructs its objects. Instead of points and lines drawn in the sand, analytic expressions—i.e., symbolic scribbles on a piece of paper—are the new primitive objects of mathematics.

In his *Théorie des fonctions analytiques* (1797), Lagrange gives a complete treatment of the calculus from this point of view. Thus he states upfront:

> It will be seen in this work that the analysis that is commonly called *transcendental* or *infinitesimal* is at bottom nothing but the analysis of primitive and derived functions, and that the differential and integral calculus is nothing, properly speaking, but the calculus of these same functions.

In other words, the ontology of mathematics simply *is* analytic expressions and nothing more. The entire framework of mathematical meaning and rigour stemming from constructions has simply been ripped away like a band-aid. The mathematical appeal of such a move is not hard to appreciate, but it comes at a cost. The framework of constructions had given mathematical concepts a clear meaning, existential status, and bond to reality. Analytical formulae have none of these things. They are scribbles on a piece of paper. Mathematics is ostensibly an empty game of symbols. One can see from its fruits that it is not empty after all, but with Lagrange mathematics has given up its attempts at explaining why.

Lagrange, thus, was determined to sever the geometrical leg of mathematics completely and mercilessly, and installing the analytical aspect—once a mere deputy in the service of geometry—in its place as the absolute ruling force of mathematics. In retrospect it is easy to see that this was a *coup d'état* a hundred years in the making. The classical geometrical paradigm could only live off past glory for so long; though once thought destined for great conquests, its attempts to stay relevant at the battlefronts of current research were becoming increasingly strained. Meanwhile, its analytical deputy was growing up fast, proving itself remarkably powerful in ways that no one could have anticipated. Soon enough it had accumulated a track record rivalling that of the geometrical paradigm in days of old. The conclusion was plain for all to see: the geometrical paradigm was not the one and only divine force in the empire of knowledge after all, but merely a passing dynasty whose cycle of power had come and gone.

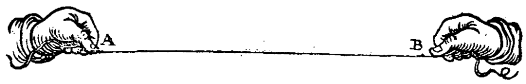
§ 23. Non-Euclidean geometry

In the early 19th century a dramatic discovery was made that profoundly changed our conception of what mathematical knowledge is and how it relates to the physical world. This discovery was that of non-Euclidean geometry: a geometry in which some but not all of Euclid's postulates hold (§5). This shook the assumption, which had been taken for granted for millennia, that Euclid's geometry is the geometry of the space in which we live.

We recall Euclid's postulates from §5. We also note a variant of Postulate 1 which Euclid and others seem to have had in mind sometimes:

> Postulate 1'. There is a *unique* straight line from any point to any point.

The question of whether this and the other postulates hold clearly depends on the meaning of "straight line." What is straightness? Euclid's definition is notoriously vague (§5). Already in Greek times Archimedes gave a better definition: A straight line is the shortest distance between two points. A more physical way of putting it would be: A straight line is the path of a stretched string. To get to the bottom of the notion of straightness it is useful to consider not only the usual plane but also other surfaces.



23.1. Consider a sphere. Answer the following questions based on the stretched-string definition of straightness.

   (a) Argue that the "straight lines" on a sphere are parts of "great circles," i.e., circles whose midpoint is the center of the sphere. The equator is an example of a great circle on the sphere of the earth.

   (b) What are some real-world applications of this fact?

   (c) Are latitude and longitude lines "straight"?

23.2. What are the straight lines on a cylinder?

23.3. Would anything change if we used Archimedes's definition instead? What does stretching a string mean in terms of distances?

To appreciate the geometry of a surface—its intrinsic geometry, as we say—we should forget for a moment that it is located in three-dimensional space. We should look at it through the eyes of a little bug who crawls around on it and thinks about its geometry but who cannot leave and is unaware of any other space beyond this surface. Think of for example those little water striders that you see running across the surfaces of ponds. They know the surface of the pond ever so well. They can feel any little movement on it. But they are quite oblivious to the existence of a third dimension outside of their surface world. This makes the water strider an easy prey for a bird or a fish that strikes it without first upsetting the surface of the water.

Thinking about the intrinsic geometry of surfaces in this way forces us to realise that what we often take for granted as "obvious" objective truths in geometry are really a lot more specific to our mental constitution and unconscious assumptions than we realise. In some ways we are as ignorant of our own limitations as the water strider.

23.4. Which of Euclid's postulates 1–5 and 1' hold on a cylinder? (Hint: you may find it useful to consider the analogy with the plane that comes from "unrolling" the cylinder into a flat plane.)

23.5. Which of Euclid's postulates 1–5 and 1' hold on a sphere?

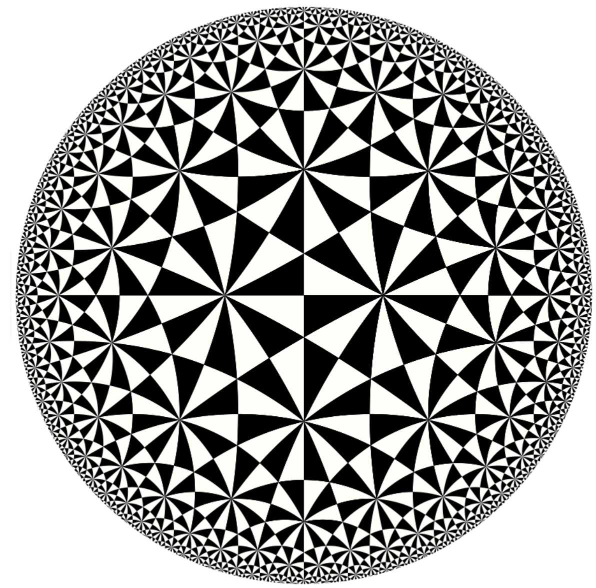


Figure 26: The conformal disc model of hyperbolic geometry.

Figure 26 is a representation of the so-called hyperbolic plane. Intrinsically, all the triangular tiles are of equal size. Thus the space represented in this picture is actually infinite since the tiles become smaller and smaller as you approach the boundary. A bug living in this world could never walk to the boundary; there would always be more tiles to go. The lines in this world are arcs of circles perpendicular to the boundary (including the diameters of the disc, which are part of circles with infinite radius, so to speak). Hyperbolic angles are the same as the Euclidean angles in the picture.

23.6. Argue that lines in this sense do indeed seem to represent the shortest (hyperbolic) distances between any two points.

23.7. Which of Euclid's postulates 1–5 and 1' are true in the hyperbolic plane?

23.8. What does this mean for the feasibility of proving the parallel postulate from the other axioms?

23.9. In what sense do the sphere, cylinder, and hyperbolic plane "prove Euclid wrong"? Why were the sphere and cylinder not considered "counterexamples" to Euclid's geometry, while the hyperbolic plane was?

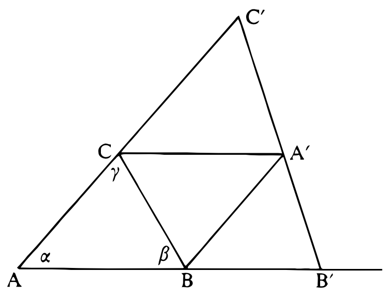
23.10. Argue that the geometry of the hyperbolic plane is "locally Euclidean," i.e., practically indistinguishable from Euclidean geometry when one zooms in far enough.

23.11. Show that there are multiple parallels to a given line through a given point in the hyperbolic plane.

23.12. Argue that the angle sum of a hyperbolic triangle is less than 180°.

23.13. Do we live in a Euclidean or hyperbolic world? How can you tell?

23.14. Legendre tried to prove that the angle sum of a triangle cannot be less than 180° using only the first four postulates of Euclid. His proof goes as follows:



Let ABC be a triangle with angles α, β, γ such that $\alpha + \beta + \gamma < 180°$. Call $180° - (\alpha + \beta + \gamma) = \delta$ the defect of the triangle. Locate A' symmetrically situated to A with respect to BC (this can be done by rotating ABC through 180° around the midpoint of BC) and extend AB and AC. Draw through A' a line meeting AB at B' and AC at C'—not necessarily a line 'parallel' to BC which would beg the question. Join up A'B and A'C. By symmetry the defect of triangle A'BC is also δ. Since the defect of the angle sum of the large triangle AB'C' is the sum of the defects in the angle sum of the four triangles separately, we have in AB'C' a triangle with defect $\geq 2\delta$. Continuing in this manner we obtain triangles with defects $\delta, 2\delta, 4\delta, 8\delta$ and so on. Eventually, then, we will have a triangle whose defect is greater than 180°, which is absurd since this would mean that $\alpha + \beta + \gamma$ is a negative number.

We have thus showed that if there is a triangle with angle sum < 180°, then there are also triangles with angle sum less than any given number, including even 0°. Therefore there can be no triangle with angle sum < 180°.

Why does Legendre's proof not work in the hyperbolic plane?

23.15. Lagrange tried to prove the uniqueness of parallels from the principle that if there are two parallels to a given line through a given point, then the reflection of one of these parallels in the other should also be a parallel of the given line (since there is no reason why one side of a line should be "privileged" over the other).

(a) Argue that Lagrange's principle implies the uniqueness of parallels.

(b) Explain why Lagrange's proof does not work in the hyperbolic plane.