PEARSON NEW INTERNATIONAL EDITION

History of Mathematics

Victor J. Katz

Third Edition

ALWAYS LEARNING™

PEARSON®

Pearson New International Edition

History of Mathematics

Victor J. Katz
Third Edition

PEARSON®

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsoned.co.uk

PEARSON®

ISBN 10: 1-292-02778-9
ISBN 13: 978-1-292-02778-4

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Printed in the United States of America

Table of Contents

CHAPTER 1

PART ONE *Ancient Mathematics*

Egypt and Mesopotamia

Accurate reckoning. The entrance into the knowledge of all existing things and all obscure secrets.

—Introduction to *Rhind Mathematical Papyrus*[1]

Mesopotamia: In a scribal school in Larsa some 3800 years ago, a teacher is trying to develop mathematics problems to assign to his students so they can practice the ideas just introduced on the relationship among the sides of a right triangle. The teacher not only wants the computations to be difficult enough to show him who really understands the material but also wants the answers to come out as whole numbers so the students will not be frustrated. After playing for several hours with the few triples (a, b, c) of numbers he knows that satisfy $a^2 + b^2 = c^2$, a new idea occurs to him. With a few deft strokes of his stylus, he quickly does some calculations on a moist clay tablet and convinces himself that he has discovered how to generate as many of these triples as necessary. After organizing his thoughts a bit longer, he takes a fresh tablet and carefully records a table listing not only 15 such triples but also a brief indication of some of the preliminary calculations. He does not, however, record the details of his new method. Those will be saved for his lecture to his colleagues. They will then be forced to acknowledge his abilities, and his reputation as one of the best teachers of mathematics will spread throughout the entire kingdom.

The opening quotation from one of the few documentary sources on Egyptian mathematics and the fictional story of the Mesopotamian scribe illustrate some of the difficulties in giving an accurate picture of ancient mathematics. Mathematics certainly existed in virtually every ancient civilization of which there are records. But in every one of these civilizations, mathematics was in the domain of specially trained priests and scribes, government officials whose job it was to develop and use mathematics for the benefit of that government in such areas as tax collection, measurement, building, trade, calendar making, and ritual practices. Yet, even though the origins of many mathematical concepts stem from their usefulness in these contexts, mathematicians always exercised their curiosity by extending these ideas far beyond the limits of practical necessity. Nevertheless, because mathematics was a tool of power, its methods were passed on only to the privileged few, often through an oral tradition. Hence, the written records are generally sparse and seldom provide much detail.

In recent years, however, a great deal of scholarly effort has gone into reconstructing the mathematics of ancient civilizations from whatever clues can be found. Naturally, all scholars do not agree on every point, but there is enough agreement so that a reasonable picture can be presented of the mathematical knowledge of the ancient civilizations in Mesopotamia and Egypt. We begin our discussion of the mathematics of each of these civilizations with a brief survey of the underlying civilization and a description of the sources from which our knowledge of the mathematics is derived.

1.1 EGYPT

Agriculture emerged in the Nile Valley in Egypt close to 7000 years ago, but the first dynasty to rule both Upper Egypt (the river valley) and Lower Egypt (the delta) dates from about 3100 BCE. The legacy of the first pharaohs included an elite of officials and priests, a luxurious court, and for the kings themselves, a role as intermediary between mortals and gods. This role fostered the development of Egypt's monumental architecture, including the pyramids, built as royal tombs, and the great temples at Luxor and Karnak. Writing began in Egypt at about this time, and much of the earliest writing concerned accounting, primarily of various types of goods. There were several different systems of measuring, depending on the particular goods being measured. But since there were only a limited number of signs, the same signs meant different things in connection with different measuring systems. From the beginning of Egyptian writing, there were two styles, the hieroglyphic writing for monumental inscriptions and the hieratic, or cursive, writing, done with a brush and ink on papyrus. Greek domination of Egypt in the centuries surrounding the beginning of our era was responsible for the disappearance of both of these native Egyptian writing forms. Fortunately, Jean Champollion (1790–1832) was able to begin the process of understanding Egyptian writing early in the nineteenth century through the help of a multilingual inscription—the Rosetta stone—in hieroglyphics and Greek as well as the later demotic writing, a form of the hieratic writing of the papyri (Fig. 1.1).

FIGURE 1.1

Jean Champollion and a piece of the Rosetta stone

It was the scribes who fostered the development of the mathematical techniques. These government officials were crucial to ensuring the collection and distribution of goods, thus helping to provide the material basis for the pharaohs' rule (Fig. 1.2). Thus, evidence for the techniques comes from the education and daily work of the scribes, particularly as related in

FIGURE 1.2

Amenhotep, an Egyptian high official and scribe (fifteenth century BCE)

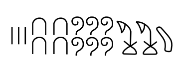
two papyri containing collections of mathematical problems with their solutions, the *Rhind Mathematical Papyrus*, named for the Scotsman A. H. Rhind (1833–1863) who purchased it at Luxor in 1858, and the *Moscow Mathematical Papyrus*, purchased in 1893 by V. S. Golenishchev (d. 1947) who later sold it to the Moscow Museum of Fine Arts. The former papyrus was copied about 1650 BCE by the scribe A'h-mose from an original about 200 years older and is approximately 18 feet long and 13 inches high. The latter papyrus dates from roughly the same period and is over 15 feet long, but only some 3 inches high. Unfortunately, although a good many papyri have survived the ages due to the generally dry Egyptian climate, it is the case that papyrus is very fragile. Thus, besides the two papyri mentioned, only a few short fragments of other original Egyptian mathematical papyri are still extant.

These two mathematical texts inform us first of all about the types of problems that needed to be solved. The majority of problems were concerned with topics involving the administration of the state. That scribes were occupied with such tasks is shown by illustrations found on the walls of private tombs. Very often, in tombs of high officials, scribes are depicted working together, probably in accounting for cattle or produce. Similarly, there exist three-dimensional models representing such scenes as the filling of granaries, and these scenes always include a scribe to record quantities. Thus, it is clear that Egyptian mathematics was developed and practiced in this practical context.

One other area in which mathematics played an important role was architecture. Numerous remains of buildings demonstrate that mathematical techniques were used both in their design and construction. Unfortunately, there are few detailed accounts of exactly how the mathematics was used in building, so we can only speculate about many of the details. We deal with a few of these ideas below.

1.1.1 Number Systems and Computations

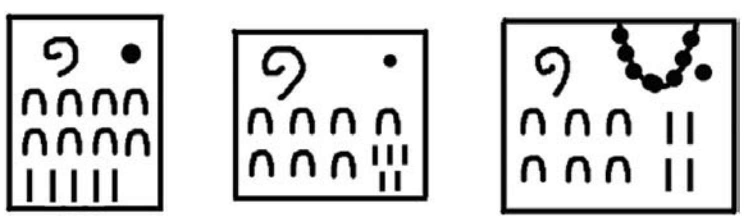
The Egyptians developed two different number systems, one for each of their two writing styles. In the hieroglyphic system, each of the first several powers of 10 was represented by a different symbol, beginning with the familiar vertical stroke for 1. Thus, 10 was represented by 𓎆, 100 by 𓍢, 1000 by 𓆼, and 10,000 by 𓂭 (Fig. 1.3). Arbitrary whole numbers were then represented by appropriate repetitions of the symbols. For example, to represent 12,643 the Egyptians would write

.

(Note that the usual practice was to put the smaller digits on the left.)

FIGURE 1.3

Egyptian numerals on the Naqada tablets (c. 3000 BCE)



The hieratic system, in contrast to the hieroglyphic, is an example of a ciphered system. Here each number from 1 to 9 had a specific symbol, as did each multiple of 10 from 10 to 90 and each multiple of 100 from 100 to 900, and so on. A given number, for example, 37, was written by putting the symbol for 7 next to that for 30. Since the symbol for 7 was and that for 30 was , 37 was written . Again, since 3 was written as , 40 as , and 200 as , the symbol for 243 was . Although a zero symbol is not necessary in a ciphered system, the Egyptians did have such a symbol. This symbol does not occur in the mathematical papyri, however, but in papyri dealing with architecture, where it is used to denote the bottom leveling line in the construction of a pyramid, and accounting, where it is used in balance sheets to indicate that the disbursements and income are equal.[2]

Once there is a system of writing numbers, it is only natural that a civilization devise algorithms for computation with these numbers. For example, in Egyptian hieroglyphics, addition and subtraction are quite simple: combine the units, then the tens, then the hundreds, and so on. Whenever a group of ten of one type of symbol appears, replace it by one of the next. Hence, to add 783 and 275,

put 𓏺𓏺𓏺𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓍢𓍢𓍢𓍢𓍢𓍢𓍢 and 𓏺𓏺𓏺𓏺𓏺𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓍢𓍢 together to get 𓏺𓏺𓏺𓏺𓏺𓏺𓏺𓏺𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓎆𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢𓍢 .

Since there are fifteen ∩'s, replace ten of them by one 𓍢. This then gives ten of the latter. Replace these by one 𓆼. The final answer is

𓏺𓏺𓏺𓏺𓏺𓏺𓏺𓏺𓎆𓎆𓎆𓎆𓎆𓆼 ,

or 1058. Subtraction is done similarly. Whenever "borrowing" is needed, one of the symbols would be converted to ten of the next lower symbol. Such a simple algorithm for addition and subtraction is not possible in the hieratic system. Probably, the scribes simply memorized basic addition tables.

The Egyptian algorithm for multiplication was based on a continual doubling process. To multiply two numbers a and b, the scribe would first write down the pair 1, b. He would then double each number in the pair repeatedly, until the next doubling would cause the first element of the pair to exceed a. Then, having determined the powers of 2 that add to a, the scribe would add the corresponding multiples of b to get his answer. For example, to multiply 12 by 13, the scribe would set down the following lines:

ˋ1	12
2	24
ˋ4	48
ˋ8	96

At this point he would stop because the next doubling would give him 16 in the first column, which is larger than 13. He would then check off those multipliers that added to 13, namely, 1, 4, and 8, and add the corresponding numbers in the other column. The result would be written as follows: Totals 13 156.

There is no record of how the scribe did the doubling. The answers are simply written down. Perhaps the scribe had memorized an extensive two times table. In fact, there is some evidence that doubling was a standard method of computation in areas of Africa to the south of Egypt and that therefore the Egyptian scribes learned from their southern colleagues. In addition, the scribes were somehow aware that every positive integer could be uniquely expressed as the sum of powers of two. That fact provides the justification for the procedure. How was it discovered? The best guess is that it was discovered by experimentation and then passed down as tradition.

Because division is the inverse of multiplication, a problem such as $156 \div 12$ would be stated as, "multiply 12 so as to get 156." The scribe would then write down the same lines as above. This time, however, he would check off the lines having the numbers in the right-hand column that sum to 156; here that would be 12, 48, and 96. Then the sum of the corresponding numbers on the left, namely, 1, 4, and 8, would give the answer 13. Of course, division does not always "come out even." When it did not, the Egyptians resorted to fractions.

The Egyptians only dealt with unit fractions or "parts" (fractions with numerator 1), with the single exception of 2/3, perhaps because these fractions are the most "natural." The fraction $1/n$ (the nth part) is in general represented in hieroglyphics by the symbol for the integer n with the symbol 𓂋 above. In the hieratic a dot is used instead. So 1/7 is denoted in the former system by 𓂋𓐀 and in the latter by $\dot{𝓏}$. The single exception, 2/3, had a special symbol: 𓂚 in hieroglyphic and ϒ in hieratic. Two other fractions, 1/2 and 1/4, also had special symbols: ⊂ and x, respectively. In what follows, however, the notation $\overline{n}$ will be used to represent $1/n$ and $\overline{\overline{3}}$ to represent 2/3.

Because fractions show up as the result of divisions that do not come out evenly, surely there is a need to be able to deal with fractions other than unit fractions. It was in this connection that the most intricate of the Egyptian arithmetical techniques developed, the representation of any fraction in terms of unit fractions. The Egyptians did not view the question this way, however. Whenever we would use a nonunit fraction, they simply wrote a sum of unit fractions. For example, problem 3 of the *Rhind Mathematical Papyrus* asks how to divide 6 loaves among 10 men. The answer is given that each man gets $\overline{2}\ \overline{10}$ loaves (that is, 1/2 + 1/10). The scribe checks this by multiplying this value by 10. We may regard the scribe's answer as more cumbersome than our answer of 3/5, but in some sense the actual division is easier to accomplish this way. We divide five of the loaves in half, the sixth one in tenths, and then give each man one half plus one tenth. It is then clear to all that every man has the same portion of bread. Cumbersome or not, this Egyptian method of unit fractions was used throughout the Mediterranean basin for over 2000 years.

In multiplying whole numbers, the important step is the doubling step. So too in multiplying fractions; the scribe had to be able to express the double of any unit fraction. For example, in the problem above, the check of the solution is written as follows:

1	$\overline{2}\ \overline{10}$
'2	$1\ \overline{5}$
4	$2\ \overline{3}\ \overline{15}$
'8	$4\ \overline{\overline{3}}\ \overline{10}\ \overline{30}$
10	6

How are these doubles formed? To double $\overline{2}\ \overline{10}$ is easy; because each denominator is even, each is merely halved. In the next line, however, $\overline{5}$ must be doubled. It was here that the scribe had to use a table to get the answer $\overline{3}\ \overline{15}$ (that is, $2 \cdot 1/5 = 1/3 + 1/15$). In fact, the first section of the *Rhind Papyrus* is a table of the division of 2 by every odd integer from 3 to 101 (Fig. 1.4), and the Egyptian scribes realized that the result of multiplying $\overline{n}$ by 2 is the same as that of dividing 2 by n. It is not known how the division table was constructed, but there are several scholarly accounts giving hypotheses for the scribes' methods. In any case, the solution of problem 3 depends on using that table twice, first as already indicated and second in the next step, where the double of $\overline{15}$ is given as $\overline{10}\ \overline{30}$ (or $2 \cdot 1/15 = 1/10 + 1/30$). The final step in this problem involves the addition of $1\ \overline{5}$ to $4\ \overline{\overline{3}}\ \overline{10}\ \overline{30}$, and here the scribe just gave the answer. Again, the conjecture is that for such addition problems an extensive table existed. The *Egyptian Mathematical Leather Roll*, which dates from about 1600 BCE, contains a short version of such an addition table.[3] There are also extant several other tables for dealing with unit fractions and a multiplication table for the special fraction 2/3. It thus appears that the arithmetic algorithms used by the Egyptian scribes involved extensive knowledge of

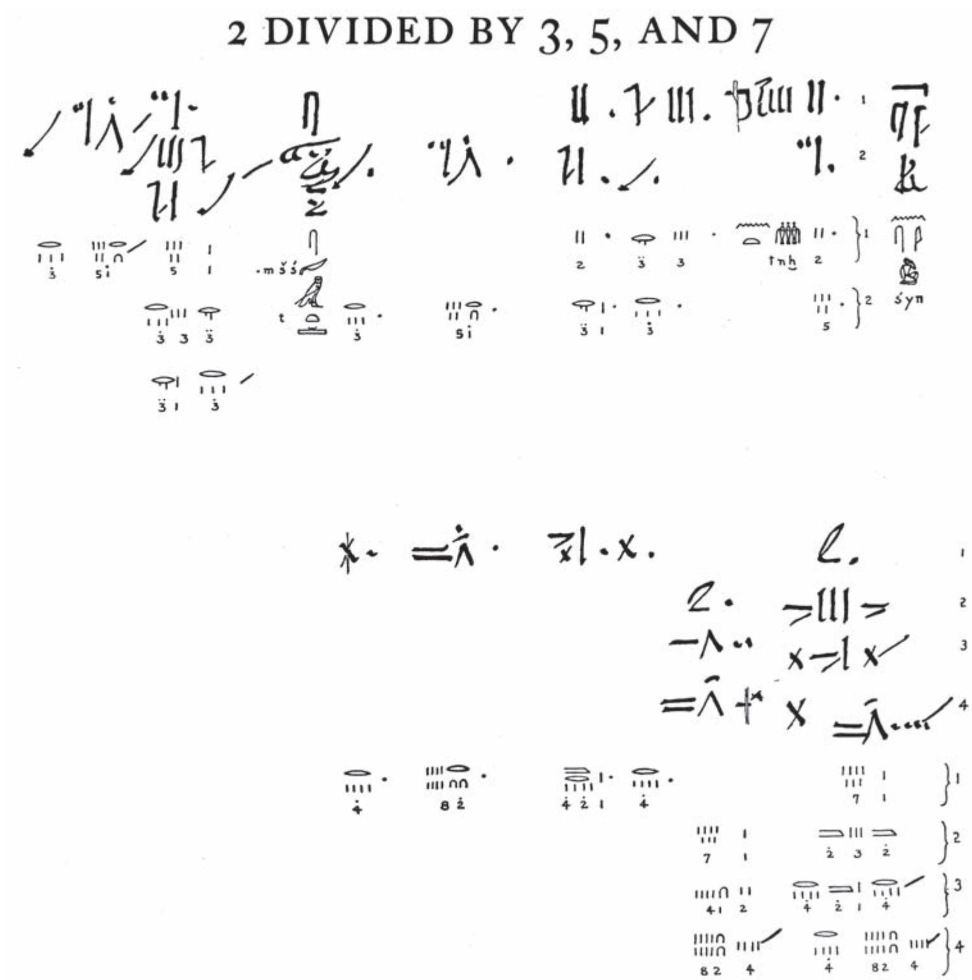


FIGURE 1.4

Transcription and hieroglyphic translation of $2 \div 3$, $2 \div 5$, and $2 \div 7$ from the *Rhind Mathematical Papyrus* (Reston, VA: National Council of Teachers of Mathematics, 1967, Arnold B. Chace, ed.)

basic tables for addition, subtraction, and doubling and then a definite procedure for reducing multiplication and division problems into steps, each of which could be done using the tables.

Besides the basic procedures of doubling, the Egyptian scribes used other techniques in performing arithmetic calculations. For example, they could find halves of numbers as well as multiply by 10; they could figure out what fractions had to be added to a given mixed number to get the next whole number; and they could determine by what fraction a given whole number needs to be multiplied to give a given fraction. These procedures are illustrated in problem 69 of the *Rhind Papyrus*, which includes the division of 80 by $3\ \overline{2}$ and its subsequent check:

1	$3\ \overline{2}$	$^{\backprime}1$	$22\ \overline{\overline{3}}\ \overline{7}\ \overline{21}$
10	35	$^{\backprime}2$	$45\ \overline{3}\ \overline{4}\ \overline{14}\ \overline{28}\ \overline{42}$
20	$70'$	$\overline{{}^{\backprime}2}$	$\underline{11\ \overline{3}\ \overline{14}\ \overline{42}}$
2	$7'$	$3\ \overline{2}$	80
$\overline{\overline{3}}$	$2\ \overline{3}'$		
$\overline{21}$	$\overline{6}'$		
$\underline{\overline{7}}$	$\underline{\overline{2}'}$		
$22\ \overline{\overline{3}}\ \overline{7}\ \overline{21}$	80		

In the second line, the scribe took advantage of the decimal nature of his notation to give immediately the product of $3\ \overline{2}$ by 10. In the fifth line, he used the 2/3 multiplication table mentioned earlier. The scribe then realized that since the numbers in the second column of the third through the fifth lines added to $79\ \overline{3}$, he needed to add $\overline{2}$ and $\overline{6}$ in that column to get 80. Thus, because $6 \times 3\ \overline{2} = 21$ and $2 \times 3\ \overline{2} = 7$, it follows that $\overline{21} \times 3\ \overline{2} = \overline{6}$ and $\overline{7} \times 3\ \overline{2} = \overline{2}$, as indicated in the sixth and seventh lines. The check shows several uses of the table of division by 2 as well as great facility in addition.

1.1.2 Linear Equations and Proportional Reasoning

The mathematical problems the scribes could solve, as illustrated in the *Rhind* and *Moscow Papyri,* deal with what we today call linear equations, proportions, and geometry. For example, the Egyptian papyri present two different procedures for dealing with linear equations.

First, problem 19 of the *Moscow Papyrus* used our normal technique to find the number such that if it is taken 1 1/2 times and then 4 is added, the sum is 10. In modern notation, the equation is simply $(1\ 1/2)x + 4 = 10$. The scribe proceeded as follows: "Calculate the excess of this 10 over 4. The result is 6. You operate on 1 1/2 to find 1. The result is 2/3. You take 2/3 of this 6. The result is 4. Behold, 4 says it. You will find that this is correct."[4] Namely, after subtracting 4, the scribe noted that the reciprocal of 1 1/2 is 2/3 and then multiplies 6 by this quantity. Similarly, problem 35 of the *Rhind Papyrus* asked to find the size of a scoop that requires 3 1/3 trips to fill a 1 hekat measure. The scribe solved the equation, which would today be written as $(3\ 1/3)x = 1$ by dividing 1 by 3 1/3. He wrote the answer as $\overline{5}\ \overline{10}$ and proceeded to prove that the result is correct.

The Egyptians' more common technique of solving a linear equation, however, was what is usually called the method of **false position**, the method of assuming a convenient but probably incorrect answer and then adjusting it by using proportionality. For example, problem 26 of the *Rhind Papyrus* asked to find a quantity such that when it is added to 1/4 of itself the result is 15. The scribe's solution was as follows: "Assume [the answer is] 4. Then 1 $\bar{4}$ of 4 is 5. . . . Multiply 5 so as to get 15. The answer is 3. Multiply 3 by 4. The answer is 12."[5] In modern notation, the problem is to solve $x + (1/4)x = 15$. The first guess is 4, because 1/4 of 4 is an integer. But then the scribe noted that $4 + 1/4 \cdot 4 = 5$. To find the correct answer, he therefore multiplied 4 by the ratio of 15 to 5, namely, 3. The *Rhind Papyrus* has several similar problems, all solved using false position. The step-by-step procedure of the scribe can therefore be considered as an algorithm for the solution of a linear equation of this type. There is, however, no discussion of how the algorithm was discovered or why it works. But it is evident that the Egyptian scribes understood the basic idea of proportionality of two quantities.

This understanding is further exemplified in the solution of more explicit proportion problems. For example, problem 75 asked for the number of loaves of *pesu* 30 that can be made from the same amount of flour as 155 loaves of *pesu* 20. (*Pesu* is the Egyptian measure for the inverse "strength" of bread and can be expressed as *pesu* = [number of loaves]/[number of hekats of grain], where a hekat is a dry measure approximately equal to 1/8 bushel.) The problem was thus to solve the proportion $x/30 = 155/20$. The scribe accomplished this by dividing 155 by 20 and multiplying the result by 30 to get 232 1/2. Similar problems occur elsewhere in the *Rhind Papyrus* and in the *Moscow Papyrus*.

On the other hand, the method of false position is also used in the only quadratic equation extant in the Egyptian papyri. On the *Berlin Papyrus*, a small fragment dating from approximately the same time as the other papyri, is a problem asking to divide a square area of 100 square cubits into two other squares, where the ratio of the sides of the two squares is 1 to 3/4. The scribe began by assuming that in fact the sides of the two needed squares are 1 and 3/4, then calculated the sum of the areas of these two squares to be $1^2 + (3/4)^2 = 1\ 9/16$. But the desired sum of the areas is 100. The scribe realized that he could not compare areas directly but must compare their sides. So he took the square root of $1\ 9/16$, namely, $1\frac{1}{4}$, and compared this to the square root of 100, namely, 10. Since 10 is 8 times as large as $1\frac{1}{4}$, the scribe concluded that the sides of the two other squares must be 8 times the original guesses, namely, 8 and 6 cubits, respectively.

There are numerous more complicated problems in the extant papyri. For example, problem 64 of the *Rhind Papyrus* reads as follows: "If it is said to thee, divide 10 hekats of barley among 10 men so that the difference of each man and his neighbor in hekats of barley is 1/8, what is each man's share?"[6] It is understood in this problem, as in similar problems elsewhere in the papyrus, that the shares are to be in arithmetic progression. The average share is 1 hekat. The largest share could be found by adding 1/8 to this average share half the number of times as there are differences. However, since there is an odd number (9) of differences, the scribe instead added half of the common difference (1/16) a total of 9 times to get $1\frac{9}{16}$ (1 $\bar{2}$ $\overline{16}$) as the largest share. He finished the problem by subtracting 1/8 from this value 9 times to get each share.

A final problem, problem 23 of the *Moscow Papyrus*, is what we often think of today as a "work" problem: "Regarding the work of a shoemaker, if he is cutting out only, he can do 10 pairs of sandals per day; but if he is decorating, he can do 5 per day. As for the number he can both cut and decorate in a day, what will that be?"[7] Here the scribe noted that the shoemaker cuts 10 pairs of sandals in one day and decorates 10 pairs of sandals in two days, so that it takes three days for him to both cut and decorate 10 pairs. The scribe then divided 10 by 3 to find that the shoemaker can cut and decorate 3 1/3 pairs in one day.

1.1.3 Geometry

As to geometry, the Egyptian scribes certainly knew how to calculate the areas of rectangles, triangles, and trapezoids by our normal methods. It is their calculation of the area of a circle, however, that is particularly interesting. Problem 50 of the *Rhind Papyrus* reads, "Example of a round field of diameter 9. What is the area? Take away 1/9 of the diameter; the remainder is 8. Multiply 8 times 8; it makes 64. Therefore, the area is 64."[8] In other words, the Egyptian scribe was using a procedure described by the formula $A = (d - d/9)^2 = [(8/9)d]^2$. A comparison with the formula $A = (\pi/4)d^2$ shows that the Egyptian value for the constant π in the case of area was $256/81 = 3.16049\ldots$. Where did the Egyptians get this value, and why was the answer expressed as the square of $(8/9)d$ rather than in modern terms as a multiple (here 64/81) of the square of the diameter?

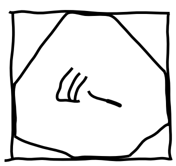


FIGURE 1.5

Octagon inscribed in a square of side 9, from problem 48 of the *Rhind Mathematical Papyrus*

A hint is given by problem 48 of the same papyrus, in which is shown the figure of an octagon inscribed in a square of side 9 (Fig. 1.5). There is no statement of the problem, however, only a bare computation of $8 \times 8 = 64$ and $9 \times 9 = 81$. If the scribe had inscribed a circle in the same square, he would have seen that its area was approximately that of the octagon. What is the size of the octagon? It depends on how one interprets the diagram in the papyrus. If one believes the octagon to be formed by cutting off four corner triangles each having area 4 1/2, then the area of the octagon is 7/9 that of the square, namely, 63. The scribe therefore might have simply taken the area of the circle as $A = (7/9)d^2[= (63/81)d^2]$. But since he wanted to find a square whose area was equal to the given circle, he may have approximated 63/81 by $(8/9)^2$, thus giving the area of the circle in the form $[(8/9)d]^2$ indicated in problem 50. On the other hand, in the diagram, the octagon does not look symmetric. So perhaps the octagon was formed by cutting off from the square of side 9 two diagonally opposite corner triangles each equal to 4 1/2 and two other corner triangles each equal to 4. This octagon then has area 64, as explicitly written on the papyrus, and thus this may be the square that the scribe wanted, which was equal in area to a circle.

It should be noted that problem 50 is not an isolated problem of finding the area of a circle. In fact, there are several problems in the *Rhind Papyrus* where the scribe used the rule $V = Bh$ to calculate the volume of a cylinder where B, the area of the base, is calculated by this circle rule. The scribes also knew how to calculate the volume of a rectangular box, given its length, width, and height.

Because one of the prominent forms of building in Egypt was the pyramid, one might expect to find a formula for its volume. Unfortunately, such a formula does not appear in any extant document. The *Rhind Papyrus* does have several problems dealing with the *seked* (slope) of a pyramid; this is measured as so many horizontal units to one vertical unit rise. The workers building the pyramids, or at least their foremen, had to be aware of this value as they built. Since the *seked* is in effect the cotangent of the angle of slope of the pyramid's faces, one can easily calculate the angles given the values appearing in the problems. It is not surprising that these calculated angles closely approximate the actual angles used in the construction of the three major pyramids at Giza.

The *Moscow Papyrus*, however, does have a fascinating formula related to pyramids, namely, the formula for the volume of a truncated pyramid (problem 14): "If someone says to you: a truncated pyramid of 6 for the height by 4 on the base by 2 on the top, you are to square this 4; the result is 16. You are to double 4; the result is 8. You are to square this 2; the result is 4. You are to add the 16 and the 8 and the 4; the result is 28. You are to take 1/3 of 6; the result is 2. You are to take 28 two times; the result is 56. Behold, the volume is 56. You will find that this is correct."[9] If this algorithm is translated into a modern formula, with the length of the lower base denoted by a, that of the upper base by b, and the height by h, it gives the correct result $V = \frac{h}{3}(a^2 + ab + b^2)$. Although no papyrus gives the formula $V = \frac{1}{3}a^2h$ for a completed pyramid of square base a and height h, it is a simple matter to derive it from the given formula by simply putting $b = 0$. We therefore assume that the Egyptians were aware of this result. On the other hand, it takes a higher level of algebraic skill to derive the volume formula for the truncated pyramid from that for the complete pyramid. Still, although many ingenious suggestions involving dissection have been given, no one knows for sure how the Egyptians found their algorithm.

No one knows either how the Egyptians found their procedure for determining the surface area of a hemisphere. But they succeeded in problem 10 of the *Moscow Papyrus*: "A basket with a mouth opening of 4 1/2 in good condition, oh let me know its surface area. First, calculate 1/9 of 9, since the basket is 1/2 of an egg-shell. The result is 1. Calculate the remainder as 8. Calculate 1/9 of 8. The result is 2/3 1/6 1/18 [that is, 8/9]. Calculate the remainder from these 8 after taking away those [8/9]. The result is 7 1/9. Reckon with 7 1/9 four and one-half times. The result is 32. Behold, this is its area. You will find that it is correct."[10] Evidently, the scribe calculated the surface area S of this basket of diameter $d = 4\ 1/2$ by first taking 8/9 of $2d$, then taking 8/9 of the result, and finally multiplying by d. As a modern formula, this result would be $S = 2(\frac{8}{9}d)^2$, or, since the area A of the circular opening of this hemispherical basket is given by $A = (\frac{8}{9}d)^2$, we could rewrite this result as $S = 2A$, the correct answer. (It should be noted that there is not universal agreement that this calculation gives the area of a hemisphere. Some suggest that it gives the surface area of a half-cylinder.)

1.2 MESOPOTAMIA

The Mesopotamian civilization is perhaps a bit older than the Egyptian, having developed in the Tigris and Euphrates River valley beginning sometime in the fifth millennium BCE. Many different governments ruled this region over the centuries. Initially, there were many small city-states, but then the area was unified under a dynasty from Akkad, which lasted from approximately 2350 to 2150 BCE. Shortly thereafter, the Third Dynasty of Ur rapidly expanded until it controlled most of southern Mesopotamia. This dynasty produced a very centralized bureaucratic state. In particular, it created a large system of scribal schools to train members of the bureaucracy. Although the Ur Dynasty collapsed around 2000 BCE, the small city-states that succeeded it still demanded numerate scribes. By 1700 BCE, Hammurapi, the ruler of Babylon, one of these city-states, had expanded his rule to much of Mesopotamia and instituted a legal system to help regulate his empire (Fig. 1.6).

FIGURE 1.6

Hammurapi on a stamp of Iraq

Writing began in Mesopotamia, quite possibly in the southern city of Uruk, at about the same time as in Egypt, namely, at the end of the fourth millennium BCE. In fact, writing began there also with the needs of accountancy, of the necessity of recording and managing labor and the flow of goods. The temple, the home of the city's patron god or goddess, came to own large tracts of farming land and vast herds of sheep and goats. The scribes of the temple managed these assets to provide for the well-being of the god(dess) and his or her followers. Thus, in the temple of goddess Inana in Uruk, the scribes represented numbers on small clay slabs, using various pictograms to represent the objects that were being counted or measured. For example, five ovoids might represent five jars of oil. Or, as in the earliest known piece of school mathematics yet discovered, the scribe who wrote tablet W 19408,76[11] used three different number signs to represent lengths as he calculated the area of a field (Fig. 1.7). Small circles represented 10 rods; a large D-shaped impression represented a unit of 60 rods, whereas a small circle within a large D represented $60 \times 10 = 600$ rods. On this tablet, there are two other signs, a horizontal line representing width and a vertical line representing length. The two widths of the quadrilateral field were each $2 \times 600 = 1200$, while the two lengths were $600 + 5 \times 60 + 3 \times 10 = 930$ and $600 + 4 \times 60 + 3 \times 10 = 870$. The approximate area could then be found by a standard ancient method of multiplying the average width by the average length; that is, $A = ((w_1 + w_2)/2)((l_1 + l_2)/2)$. In this case, the answer was $1200 \times 900 = 1{,}080{,}000$. But since in the then current measurement system 1 square rod was equal to 1 *sar*, while 1800 *sar* were equal to 1 *bur*, the result here was 600 *bur*, a conspicuously "round" number, typical of answers in school tablets.

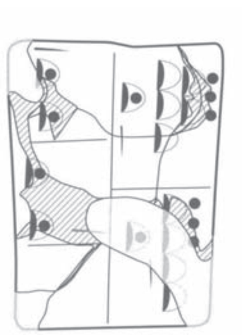


FIGURE 1.7

Tablet from Uruk, c. 3200 BCE, with number signs

On this particular tablet, as in other situations where quantities were measured, there were several different units of measure and different symbols for each type of unit. Here, the largest unit was equal to 60 of the smallest unit. This was typical in the units for many different types of objects, and at some time, the system of recording numbers developed to the point where the digit for 1 represented 60 as well. We do not know why the Mesopotamians decided to have one large unit represent 60 small units and then adapt this method for their numeration system. One plausible conjecture is that 60 is evenly divisible by many small integers. Therefore, fractional values of the "large" unit could easily be expressed as integral values of the "small." But eventually, they did develop a sexagesimal (base-60) place value system, which in the third millennium BCE became the standard system used throughout Mesopotamia. By that time, too, writing began to be used in a wide variety of contexts, all achieved by using a stylus on a moist clay tablet (Fig. 1.8). Thousands of these tablets have been excavated during the past 150 years. It was Henry Rawlinson (1810–1895) who, by the mid-1850s, was first able to translate this cuneiform writing by comparing the Persian and Mesopotamian cuneiform inscriptions of King Darius I of Persia (sixth century BCE) on a rock face at Behistun (in modern Iran) describing a military victory.

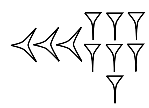
FIGURE 1.8

Babylonian clay tablet on a stamp of Austria

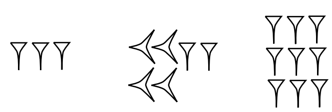
A large number of these tablets are mathematical in nature, containing mathematical problems and solutions or mathematical tables. Several hundreds of these have been copied, translated, and explained. These tablets, generally rectangular but occasionally round, usually fit comfortably into one's hand and are an inch or so in thickness. Some, however, are as small as a postage stamp while others are as large as an encyclopedia volume. We are fortunate that these tablets are virtually indestructible, because they are our only source for Mesopotamian mathematics. The written tradition that they represent died out under Greek domination in the last centuries BCE and was totally lost until the nineteenth century. The great majority of the excavated tablets date from the time of Hammurapi, while small collections date from the earliest beginnings of Mesopotamian civilization, from the centuries surrounding 1000 BCE, and from the Seleucid period around 300 BCE. Our discussion in this section, however, will generally deal with the mathematics of the "Old Babylonian" period (the time of Hammurapi), but, as is standard in the history of mathematics, we shall use the adjective "Babylonian" to refer to the civilization and culture of Mesopotamia, even though Babylon itself was the major city of the area for only a limited time.

1.2.1 Methods of Computation

The Babylonians at various times used different systems of numbers, but the standardized system that the scribes generally used for calculations in the "Old Babylonian" period was a base-60 place value system together with a grouping system based on 10 to represent numbers up to 59. Thus, a vertical stylus stroke on a clay tablet  represented 1 and a tilted stroke  represented 10. By grouping they would, for example, represent 37 by

.

For numbers greater than 59, the Babylonians used a place value system; that is, the powers of 60, the base of this system, are represented by "places" rather than symbols, while the digit in each place represents the number of each power to be counted. Hence, $3 \times 60^2 + 42 \times 60 + 9$ (or 13,329) was represented by the Babylonians as

.

(This will be written from now on as 3,42,09 rather than with the Babylonian strokes.) The Old Babylonians did not use a symbol for 0, but often left an internal space if a given number was missing a particular power. There would not be a space at the end of a number, making it difficult to distinguish $3 \times 60 + 42$ (3,42) from $3 \times 60^2 + 42 \times 60$ (3,42,00). Sometimes, however, they would give an indication of the absolute size of a number by writing an appropriate word, typically a metrological one, after the numeral. Thus, "3 42 sixty" would represent 3,42, while "3 42 thirty-six hundred" would mean 3,42,00. On the other hand, the Babylonians never used a symbol to represent zero in the context of "nothingness," as in our $42 - 42 = 0$.

That the Babylonians used tables in the process of performing arithmetic computations is proved by extensive direct evidence. Many of the preserved tablets are in fact multiplication tables. No addition tables have turned up, however. Because over 200 Babylonian table texts have been analyzed, it may be assumed that these did not exist and that the scribes knew their addition procedures well enough so they could write down the answers when needed. On the other hand, there are many examples of "scratch tablets," on which a scribe has performed various calculations in the process of solving a problem. In any case, since the Babylonian number system was a place value system, the actual algorithms for addition and subtraction, including carrying and borrowing, may well have been similar to modern ones. For example, to add 23,37 (= 1417) to 41,32 (= 2492), one first adds 37 and 32 to get 1,09 (= 69). One writes down 09 and carries 1 to the next column. Then $23 + 41 + 1 = 1, 05$ (= 65), and the final result is 1,05,09 (= 3909).

Because the place value system was based on 60, the multiplication tables were extensive. Any given one listed the multiples of a particular number, say, 9, from 1×9 to 20×9 and then gave 30×9, 40×9, and 50×9 (Fig. 1.9). If one needed the product 34×9, one simply added the two results $30 \times 9 = 4, 30$ (= 270) and $4 \times 9 = 36$ to get 5,06 (= 306). For multiplication of two- or three-digit sexagesimal numbers, one needed to use several such tables. The exact algorithm the Babylonians used for such multiplications—where the partial products are written and how the final result is obtained—is not known, but it may well have been similar to our own.

One might think that for a complete system of tables, the Babylonians would have one for each integer from 2 to 59. Such was not the case, however. In fact, although there are no tables

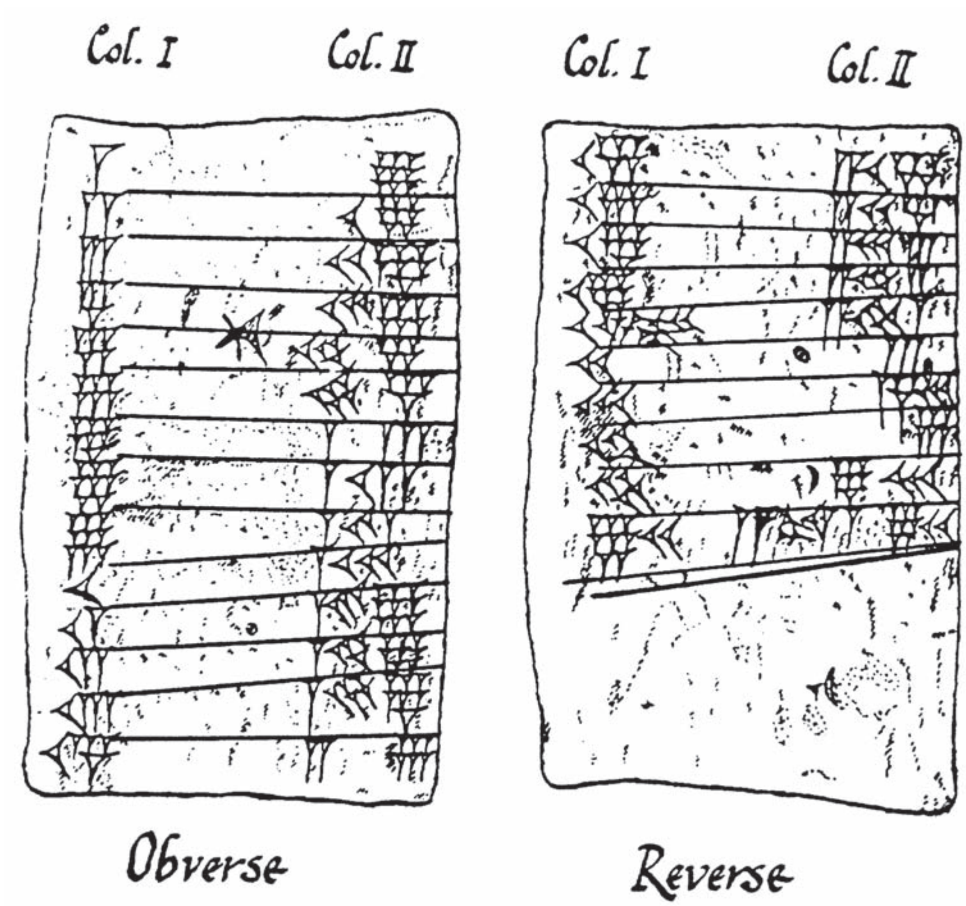


FIGURE 1.9

A Babylonian multiplication table for 9 (Department of Archaeology, University of Pennsylvania) for 11, 13, 17, for example, there are tables for 1,15, 3,45, and 44,26,40. We do not know precisely why the Babylonians made these choices; we do know, however, that, with the single exception of 7, all multiplication tables so far found are for **regular** sexagesimal numbers, that is, numbers whose reciprocal is a terminating sexagesimal fraction. The Babylonians treated all fractions as sexagesimal fractions, analogous to our use of decimal fractions. Namely, the first place after the "sexagesimal point" (which we denote by ";") represents 60ths, the next place 3600ths, and so on. Thus, the reciprocal of 48 is the sexagesimal fraction 0;1,15, which represents $1/60 + 15/60^2$, while the reciprocal of 1,21 (= 81) is 0;0,44,26,40, or $44/60^2 + 26/60^3 + 40/60^4$. Because the Babylonians did not indicate an initial 0 or the sexagesimal point, this last number would just be written as 44,26,40. As noted, there exist multiplication tables for this regular number. In such a table there is no indication of the absolute size of the number, nor is one necessary. When the Babylonians used the table, of course, they realized that, as in today's decimal calculations, the eventual placement of the sexagesimal point depended on the absolute size of the numbers involved, and this placement was then done by context.

Besides multiplication tables, there are also extensive tables of reciprocals, one of which is in part reproduced here. A table of reciprocals is a list of pairs of numbers whose product is 1 (where the 1 can represent any power of 60). Like the multiplication tables, these tables only contained regular sexagesimal numbers.

2	30	16	3, 45	48	1, 15
3	20	25	2, 24	1, 04	56, 15
10	6	40	1, 30	1, 21	44, 26, 40

The reciprocal tables were used in conjunction with the multiplication tables to do division. Thus, the multiplication table for 1,30 (= 90) served not only to give multiples of that number but also, since 40 is the reciprocal of 1,30, to do divisions by 40. In other words, the Babylonians considered the problem $50 \div 40$ to be equivalent to $50 \times 1/40$, or in sexagesimal notation, to $50 \times 0;1,30$. The multiplication table for 1,30, part of which appears here, then gives 1,15 (or 1,15,00) as the product. The appropriate placement of the sexagesimal point gives 1;15(= 1 1/4) as the correct answer to the division problem.

1	1,30	10	15	30	45
2	3	11	16,30	40	1
3	4,30	12	18	50	1,15

1.2.2 Geometry

The Babylonians had a wide range of problems to which they applied their sexagesimal place value system. For example, they developed procedures for determining areas and volumes of various kinds of figures. They worked out algorithms to determine square roots. They solved problems that we would interpret in terms of linear and quadratic equations, problems often related to agriculture or building. In fact, the mathematical tablets themselves are generally concerned with the solution of problems, to which various mathematical techniques are applied. So we will look at some of the problems the Babylonians solved and try to figure out what lies behind their methods. In particular, we will see that the reasons behind many of the Babylonian procedures come from a tradition different from the accountancy traditions with which Babylonian mathematics began. This second tradition was the "cut-and-paste" geometry of the surveyors, who had to measure fields and lay out public works projects. As we will see, these manipulations of squares and rectangles not only developed into procedures for determining square roots and finding Pythagorean triples, but they also developed into what we can think of as "algebra."

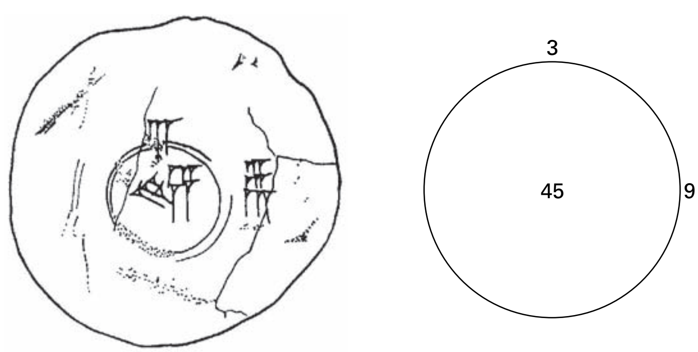
As we work through the Babylonian problems, we must keep in mind that, like the Egyptians, the scribes did not have any symbolism for operations or unknowns. Thus, solutions are presented with purely verbal techniques. We must also remember that the Babylonians often thought about problems in ways different from the ways we do. Thus, even though their methods are usually correct, they may seem strange to us.

As one example of the scribes' different methods, we consider their procedures for determining lengths and areas. In general, in place of our formulas for calculating such quantities, they presented coefficient lists, lists of constants that embody mathematical relationships between certain aspects of various geometrical figures. Thus, the number 0;52,30 (= 7/8) as the coefficient for the height of a triangle means that the altitude of an equilateral triangle is 7/8 of the base, while the number 0;26,15 ($= 7/16$) as the coefficient for area means that the area of an equilateral triangle is 7/16 times the square of a side. (Note, of course, that these results are only approximately correct, in that they both approximate $\sqrt{3}$ by 7/4.) In each case, the idea is that the "defining component" for the triangle is the side.

We too would use the length of a side as the defining component for an equilateral triangle. But for a circle, we generally use the radius r as that component and therefore give formulas for the circumference and area in terms of r. The Babylonians, on the other hand, took the circumference as the defining component of a circle. Thus, they gave two coefficients for the circle: 0;20 ($= 1/3$) for the diameter and 0;05 ($= 1/12$) for the area. The first coefficient means that the diameter is one-third of the circumference, while the second means that the area is one-twelfth of the square of the circumference. For example, on the tablet YBC 7302, there is a circle with the numbers 3 and 9 written on the outside and the number 45 written on the inside (Fig. 1.10). The interpretation of this is that the circle has circumference 3 and that the area is found by dividing $9 = 3^2$ by 12 to get 0;45 ($= 3/4$). Another tablet, Haddad 104, illustrates that circle calculations virtually always use the circumference. On this tablet, there is a problem asking to find the area of the cross section of a log of diameter 1;40 ($= 1\frac{2}{3}$). Rather than determine the radius, the scribe first multiplies by 3 to find that the circumference

FIGURE 1.10

Tablet YBC 7302 illustrating measurements on a circle



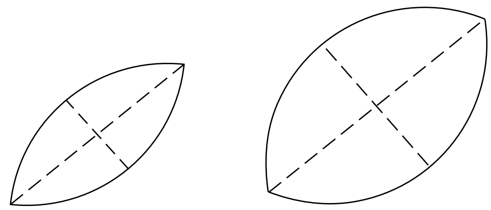


FIGURE 1.11
Babylonian barge and bull's-eye

is equal to 5, then squares 5 and multiplies by $1/12$ to get the area 2;05 ($= 2\frac{1}{12}$). Note further, of course, that the Babylonian value for what we denote as π, the ratio of circumference to diameter, is 3; this value produces the value $4\pi = 12$ as the constant by which to divide the square of the circumference to give the area.

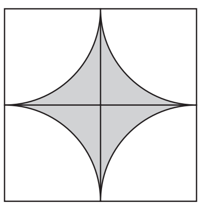


FIGURE 1.12
Babylonian concave square

There are also Babylonian coefficients for other figures bounded by circular arcs. For example, the Babylonians calculated areas of two different double bows: the "barge," made up of two quarter-circle arcs, and the "bull's-eye," composed of two third-circle arcs (Fig. 1.11). In analogy with the circle, the defining component of these figures was the arc making up one side. The coefficient of the area of the barge is 0;13,20 ($= 2/9$), while that of the bull's-eye is 0;16,52,30 ($= 9/32$). Thus, the areas of these two figures are calculated as $(2/9)a^2$ and $(9/32)a^2$, respectively, where in each case a is the length of that arc. These results are accurate under the assumptions that the area of the circle is $C^2/12$ and that $\sqrt{3} = 7/4$. Similarly, the coefficient of the area of the concave square (Fig. 1.12) is 0;26,40 ($= 4/9$), where the defining component is one of the four quarter-circle arcs forming the boundary of the region.[12] Clearly, the use of these coefficients shows that the scribes recognized that lengths of particular lines in given figures were proportional to the length of the defining component, while the area was proportional to the square of that component.

The Babylonians also dealt with volumes of solids. They realized that the volume V of a rectangular block is $V = \ell w h$, and they also knew how to calculate the volume of prisms given the area of the base. But just like in Egypt, there is no document that explicitly gives the volume of a pyramid, even though the Babylonians certainly built pyramidal structures. Nevertheless, on tablet BM 96954, there are several problems involving a grain pile in the shape of a rectangular pyramid with an elongated apex, like a pitched roof (Fig. 1.13). The method of solution corresponds to the modern formula

$$V = \frac{hw}{3}\left(\ell + \frac{t}{2}\right),$$

where ℓ is the length of the solid, w the width, h the height, and t the length of the apex. Although no derivation of this correct formula is given on the tablet, we can derive it by breaking up the solid into a triangular prism with half a rectangular pyramid on each side. Then the volume would be the sum of the volumes of these solids (Fig. 1.14). Thus, V = volume of triangular prism + volume of rectangular pyramid, or

$$V = \frac{hwt}{2} + \frac{hw(\ell - t)}{3} = \frac{hw\ell}{3} + \frac{hwt}{6} = \frac{hw}{3}\left(\ell + \frac{t}{2}\right),$$

as desired.[13] It therefore seems reasonable to assume from the result discussed here that the Babylonians were aware of the correct formula for the volume of a pyramid.

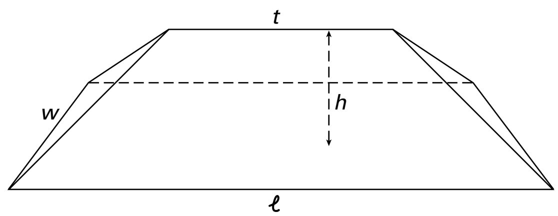


FIGURE 1.13
Babylonian grain pile

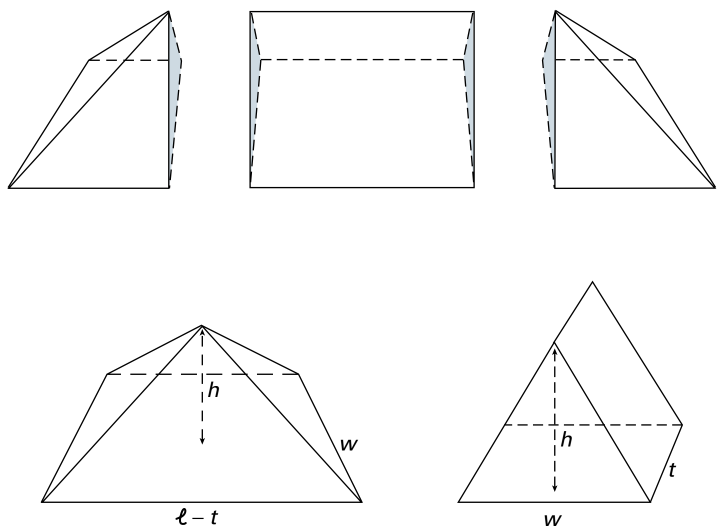


FIGURE 1.14
Dissection of grain pile

That assumption is even more convincing because there is a tablet giving a correct formula for the volume of a truncated pyramid with square base a^2, square top b^2, and height h in the form $V = [(\frac{a+b}{2})^2 + \frac{1}{3}(\frac{a-b}{2})^2]h$. The complete pyramid formula, of course, follows from this by putting $b = 0$. On the other hand, there are tablets where this volume is calculated by the rule $V = \frac{1}{2}(a^2 + b^2)h$, a simple but incorrect generalization of the rule for the area of the trapezoid. It is well to remember, however, that although this formula is incorrect, the calculated answers would not be very different from the correct ones. It is difficult to see how anyone would realize that the answers were wrong in any case, because there was no accurate method for measuring the volume empirically. However, because the problems in which these formulas occurred were practical ones, often related to the number of workmen needed to build a particular structure, the slight inaccuracy produced by using this rule would have little effect on the final answer.

1.2.3 Square Roots and the Pythagorean Theorem

We next consider another type of Babylonian algorithm, the square root algorithm. Usually, when square roots are needed in solving problems, the problems are arranged so that the square root is one that is listed in a table of square roots, of which many exist, and is a rational number. But there are cases where an irrational square root is needed, in particular,

$\sqrt{2}$. When this particular value occurs, the result is generally written as 1;25 (= $1\frac{5}{12}$). There is, however, an interesting tablet, YBC 7289, on which is drawn a square with side indicated as 30 and two numbers, 1:24,51,10 and 42;25,35, written on the diagonal (Fig. 1.15). The product of 30 by 1;24,51,10 is precisely 42;25,35. It is then a reasonable assumption that the last number represents the length of the diagonal and that the other number represents $\sqrt{2}$.

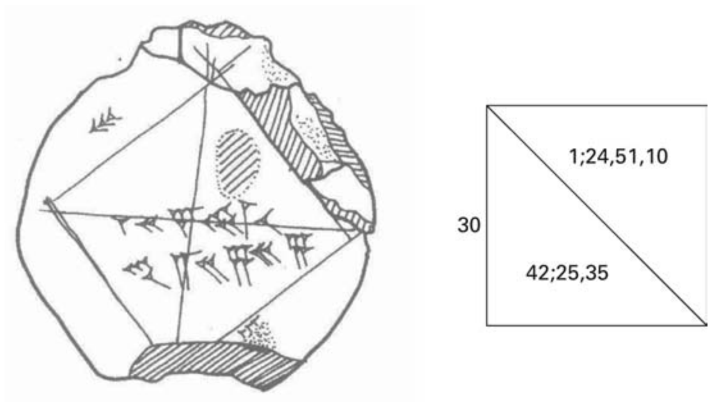


FIGURE 1.15
Tablet YBC 7289 with the square root of 2

Whether $\sqrt{2}$ is given as 1;25 or as 1;24,51,10, there is no record as to how the value was calculated. But because the scribes were surely aware that the square of neither of these was exactly 2, or that these values were not exactly the length of the side of a square of area 2, they must have known that these values were approximations. How were they determined? One possible method, a method for which there is some textual evidence, begins with the algebraic identity $(x + y)^2 = x^2 + 2xy + y^2$, whose validity was probably discovered by the Babylonians from its geometric equivalent. Now given a square of area N for which one wants the side $\sqrt{N}$, the first step would be to choose a regular value a close to, but less than, the desired result. Setting $b = N - a^2$, the next step is to find c so that $2ac + c^2$ is as close as possible to b (Fig. 1.16). If a^2 is "close enough" to N, then c^2 will be small in relation to $2ac$, so c can be chosen to equal $(1/2)b(1/a)$, that is, $\sqrt{N} = \sqrt{a^2 + b} \approx a + (1/2)b(1/a)$. (In keeping with Babylonian methods, the value for c has been written as a product rather than a quotient, and, since one of the factors is the reciprocal of a, we see why a must be regular.) A similar argument shows that $\sqrt{a^2 - b} \approx a - (1/2)b(1/a)$. In the particular case of $\sqrt{2}$, one begins with $a =$ 1;20 $(= 4/3)$. Then $a^2 =$ 1;46,40, $b =$ 0;13,20, and $1/a =$ 0;45, so $\sqrt{2} = \sqrt{1;46,40 + 0;13,20} \approx 1;20 + (0;30)(0;13,20)(0;45) = 1;20 + 0;05 = 1;25$ (or 17/12).

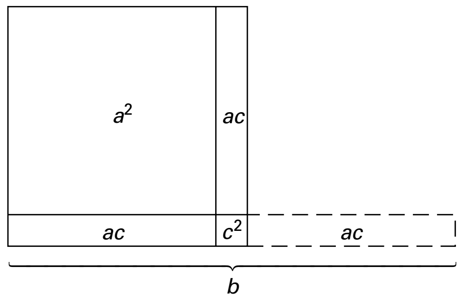
To calculate the better approximation 1;24,51,10, one would have to repeat this procedure, with $a =$ 1;25. Unfortunately, 1;25 is not a regular sexagesimal number. The scribes could, however, have found an approximation to the reciprocal, say, 0;42,21,10, and then calculated

$$\sqrt{2} = \sqrt{1;25^2 - 0;00,25} \approx 1;25 - 0;30 \times 0;00,25 \times 0;42,21,10 = 1;24,51,10,35,25.$$

Because the approximation formula leads to a slight overestimate of the true value, the scribes would have truncated this answer to the desired 1;24,51,10. There is, however, no direct

FIGURE 1.16

Geometric version of $\sqrt{N} = \sqrt{a^2+b} \approx a + \frac{1}{2} \cdot b \cdot \frac{1}{a}$



evidence of this calculation nor even any evidence for the use of more than one step of this approximation procedure.

One of the Babylonian square root problems was connected to the relation between the side of a square and its diagonal. That relation is a special case of the result known as the **Pythagorean Theorem**: In any right triangle, the sum of the areas of the squares on the legs equals the area of the square on the hypotenuse. This theorem, named after the sixth-century BCE Greek philosopher-mathematician, is arguably the most important elementary theorem in mathematics, since its consequences and generalizations have wide-ranging application. Nevertheless, it is one of the earliest theorems known to ancient civilizations. In fact, there is evidence that it was known at least 1000 years before Pythagoras.

In particular, there is substantial evidence of interest in Pythagorean triples, triples of integers (a, b, c) such that $a^2 + b^2 = c^2$, in the Babylonian tablet Plimpton 322 (Fig. 1.17).[14] The extant piece of the tablet consists of four columns of numbers. Other columns were probably broken off on the left. The numbers on the tablet are shown in Table 1.1, reproduced in modern decimal notation with the few corrections that recent editors have made and with one extra column, y (not on the tablet), added on the right. It was a major piece of mathematical detective work for modern scholars, first, to decide that this was a mathematical work rather than a list of orders from a pottery business and, second, to find a reasonable mathematical explanation. But find one they did. The columns headed x and d (whose headings in the original can be translated as "square-side of the short side" and "square-side of the diagonal") contain in each row two of the three numbers of a Pythagorean triple. It is easy enough to subtract the square of column x from the square of column d. In each case a perfect square results, whose square root is indicated in the added column, y. Finally, the first column on the left represents the quotient $(\frac{d}{y})^2$.

How and why were these triples derived? One cannot find Pythagorean triples of this size by trial and error. There have been many suggestions over the years as to how the scribe found these as well as to the purpose of the tablet. If one considers this question as purely a mathematical one, there are many methods that would work to generate the table. But since this tablet was written at a particular time and place, probably in Larsa around 1800 BCE, an understanding of its construction and meaning must come from an understanding of the context of the time and how mathematical tablets were generally written. In particular, it is important to note that the first column in a Babylonian table is virtually always written in numerical order (either ascending or descending), while subsequent columns depend on those to their left. Unfortunately, in this instance it is believed that the initial columns

FIGURE 1.17
Plimpton 322 (*Source:* George Arthur Plimpton Collection, Rare Book and Manuscript Library, Columbia University)

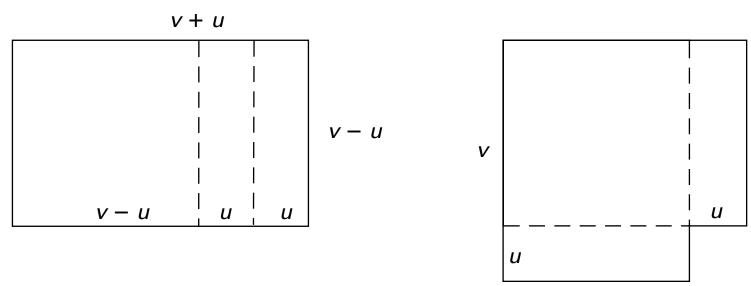
TABLE 1.1 ***Numbers on the Babylonian tablet Plimpton 322, reproduced in modern decimal notation. (The column to the right, labeled y, does not appear on the tablet.)***

$\left(\frac{d}{y}\right)^2$	x	d	#	y
1.9834028	119	169	1	120
1.9491586	3367	4825	2	3456
1.9188021	4601	6649	3	4800
1.8862479	12,709	18,541	4	13,500
1.8150077	65	97	5	72
1.7851929	319	481	6	360
1.7199837	2291	3541	7	2700
1.6845877	799	1249	8	960
1.6426694	481	769	9	600
1.5861226	4961	8161	10	6480
1.5625	45	75	11	60
1.4894168	1679	2929	12	2400
1.4500174	161	289	13	240
1.4302388	1771	3229	14	2700
1.3871605	28	53	15	45

on the left are missing. However, some clues as to the meaning of the table reside in the words at the top of the column we have labeled $(\frac{d}{y})^2$. Deciphering the words was difficult because some of the cuneiform wedges were damaged, but it appears that the heading means "the holding-square of the diagonal from which 1 is torn out so that the short side comes up." The "1" in that heading indicates that the author is dealing with reciprocal pairs, very common in Babylonian tables. To relate reciprocals to Pythagorean triples, we note that to find integer solutions to the equation $x^2 + y^2 = d^2$, one can divide by y and first find solutions to $(\frac{x}{y})^2 + 1 = (\frac{d}{y})^2$ or, setting $u = \frac{x}{y}$ and $v = \frac{d}{y}$, to $u^2 + 1 = v^2$. This latter equation is equivalent to $(v + u)(v - u) = 1$. That is, we can think of $v + u$ and $v - u$ as the sides of a rectangle whose area is 1 (Fig. 1.18). Now split off from this rectangle one with sides u and $v - u$ and move it to the bottom left after a rotation of 90°. The resulting figure is an L-shaped figure, usually called a gnomon, with long sides both equal to v, a figure that is the difference $v^2 - u^2 = 1$ of two squares. Note that the larger square is the square on the diagonal of the right triangle with sides $(u, 1, v)$. The area of that square, $v^2 = (d/y)^2$, is the entry in the leftmost column on the extant tablet, and furthermore, that square has a gnomon of area 1 torn out so that the remaining square is the square on the short side of the right triangle, as the column heading actually says.

FIGURE 1.18

A rectangle of area 1 turned into the difference of two squares



To calculate the entries on the tablet, it is possible that the author began with a value for what we have called $v + u$. Next, he found its reciprocal $v - u$ in a table and solved for $u = \frac{1}{2}[(v + u) - (v - u)]$. The first column in the table is then the value $1 + u^2$. He could then find v by taking the square root of $1 + u^2$. Since $(u, 1, v)$ satisfies the Pythagorean identity, the author could find a corresponding integral Pythagorean triple by multiplying each of these values by a suitable number y, one chosen to eliminate "fractional" values. For example, if $v + u = 2;15$ $(= 2\frac{1}{4})$, the reciprocal $v - u$ is $0;26,40$ $(= 4/9)$. We then find $u = 0;54,10 = 65/72$. We would find v by taking half the sum of $v + u$ and $v - u$, but our scribe found v as $\sqrt{1 + u^2} = \sqrt{1;48,54,01,40} = 1;20,50$, or $\sqrt{1 + u^2} = \sqrt{1.8150077} = 1\frac{25}{72}$. Multiplying the values for u, v, and 1 by $1,12 = 72$ gives the values 65 and 97 for x and d, respectively, shown in line 5 of the table, as well as the value 72 for y. Conversely, the value of $v + u$ for line 1 of the table can be found by adding $169/120$ $(= 1;24,30)$ and $119/120$ $(= 0;59,30)$ to get $288/120$ $(= 2;24)$.

Why were the particular Pythagorean triples on this tablet chosen? Again, we cannot know the answer definitively. But if we calculate the values of $v + u$ for every line of the tablet, we notice that they form a decreasing sequence of regular sexagesimal numbers of no more than four places from 2;24 to 1;48. Not all such numbers are included—there are five missing—but it is possible that the scribe may have decided that the table was long enough without them. He may also have begun with numbers larger than 2;24 or continued with numbers smaller than 1;48 on tablets that have not yet been unearthed. In any case, it is likely that this column of values for $v + u$, in descending numerical order, was one of the missing columns on the original tablet. And our author, quite probably a teacher, had thus worked out a list of integral Pythagorean triples, triples that could be used in constructing problems for students for which he would know that the solution would be possible in integers or finite sexagesimal fractions.

Whether or not the method presented above was the one the Babylonian scribe used to write Plimpton 322, the fact remains that the scribes were well aware of the Pythagorean relationship. And although this particular table offers no indication of a geometrical relationship except for the headings of the columns, there are problems in Old Babylonian tablets making explicit geometrical use of the Pythagorean Theorem. For example, in a problem from tablet BM 85196, a beam of length 30 stands against a wall. The upper end has slipped down a distance 6. How far did the lower end move? Namely, $d = 30$ and $y = 24$ are given, and x is to be found. The scribe calculated x using the theorem: $x = \sqrt{30^2 - 24^2} = \sqrt{324} = 18$. Another slightly more complicated example comes from tablet TMS 1 found at Susa in modern Iran. The problem is to calculate the radius of a circle circumscribed about an isosceles triangle with altitude 40 and base 60. Applying the Pythagorean theorem to the right triangle ABC (Fig. 1.19), whose hypotenuse is the desired radius, gives the relationship $r^2 = 30^2 + (40 - r)^2$. This could be easily transformed into $(1, 20)(r - 20) = 15{,}00$ and then, by multiplying by the reciprocal 0;0,45 of 1,20, into $r - 20 = (0;0{,}45)(15{,}00) = 11;15$, from which the scribe found that $r = 31;15$.

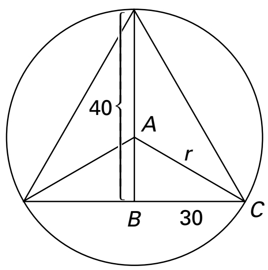


FIGURE 1.19

Circumscribing a circle about an isosceles triangle

1.2.4 Solving Equations

The previous problem involved what we would call the solution of an equation. Such problems were very frequent on the Babylonian tablets. Linear equations of the form $ax = b$ are generally solved by multiplying each side by the reciprocal of a. (Such equations often occur, as in the previous example, in the process of solving a complex problem.) In more complicated situations, such as systems of two linear equations, the Babylonians, like the Egyptians, used the method of false position.

Here is an example from the Old Babylonian text VAT 8389: One of two fields yields 2/3 *sila* per *sar*, the second yields 1/2 *sila* per *sar*, where *sila* and *sar* are measures for capacity and area, respectively. The yield of the first field was 500 *sila* more than that of the second; the areas of the two fields were together 1800 *sar*. How large is each field? It is easy enough to translate the problem into a system of two equations with x and y representing the unknown areas:

$$\frac{2}{3}x - \frac{1}{2}y = 500$$
$$x + y = 1800$$

A modern solution might be to solve the second equation for x and substitute the result in the first. But the Babylonian scribe here made the initial assumption that x and y were both equal to 900. He then calculated that $(2/3) \cdot 900 - (1/2) \cdot 900 = 150$. The difference between the desired 500 and the calculated 150 is 350. To adjust the answers, the scribe presumably realized that every unit increase in the value of x and consequent unit decrease in the value of y gave an increase in the "function" $(2/3)x - (1/2)y$ of $2/3 + 1/2 = 7/6$. He therefore needed only to solve the equation $(7/6)s = 350$ to get the necessary increase $s = 300$. Adding 300 to 900 gave him 1200 for x while subtracting gave him 600 for y, the correct answers.

Presumably, the Babylonians also solved complex single linear equations by false position, although the few such problems available do not reveal their method. For example, here is a problem from tablet YBC 4652: "I found a stone, but did not weigh it; after I added one-seventh and then one-eleventh [of the total], it weighed 1 *mina* [= 60 *gin*]. What was the original weight of the stone?"[15] We can translate this into the modern equation $(x + x/7) + 1/11(x + x/7) = 60$. On the tablet, the scribe just presented the answer, here $x = 48\frac{1}{8}$. If he had solved the problem by false position, the scribe would first have guessed that $y = x + x/7 = 11$. Since then $y + (1/11)y = 12$ instead of 60, the guess must be increased by the factor $60/12 = 5$ to the value 55. Then, to solve $x + x/7 = 55$, the scribe could have guessed $x = 7$. This value would produce $7 + 7/7 = 8$ instead of 55. So the last step would be to multiply the guess of 7 by the factor 55/8 to get $385/8 = 48\frac{1}{8}$, the correct answer.

While tablets containing explicit linear problems are limited, there are very many Babylonian tablets whose problems can be translated into quadratic equations. In fact, many Old Babylonian tablets contain extensive lists of quadratic problems. And in solving these problems, the scribes made full use of the "cut-and-paste" geometry developed by the surveyors. In particular, they applied this to various standard problems such as finding the length and width of a rectangle, given the semiperimeter and the area. For example, consider the problem $x + y = 6\frac{1}{2}$, $xy = 7\frac{1}{2}$ from tablet YBC 4663. The scribe first halved $6\frac{1}{2}$ to get $3\frac{1}{4}$. Next he squared $3\frac{1}{4}$, getting $10\frac{9}{16}$. From this is subtracted $7\frac{1}{2}$, leaving $3\frac{1}{16}$, and then the square root is extracted to get $1\frac{3}{4}$. The length is thus $3\frac{1}{4} + 1\frac{3}{4} = 5$, while the width is given as $3\frac{1}{4} - 1\frac{3}{4} = 1\frac{1}{2}$. A close reading of the wording of the tablets indicates that the scribe had in mind a geometric procedure (Fig. 1.20), where for the sake of generality the sides have been labeled in accordance with the generic system $x + y = b$, $xy = c$. The scribe began by halving the sum b and then constructing the square on it. Since $b/2 = x - \frac{x-y}{2} = y + \frac{x-y}{2}$, the square on $b/2$ exceeds the original rectangle of area c by the square on $\frac{x-y}{2}$; that is,

$$\left(\frac{x+y}{2}\right)^2 = xy + \left(\frac{x-y}{2}\right)^2.$$

The figure then shows that if one adds the side of this square, namely,

$$\sqrt{(b/2)^2 - c},$$

to $b/2$, one finds the length x, while if one subtracts it from $b/2$, one gets the width y. The algorithm is therefore expressible in the form

$$x = \frac{b}{2} + \sqrt{(b/2)^2 - c} \qquad y = \frac{b}{2} - \sqrt{(b/2)^2 - c}.$$

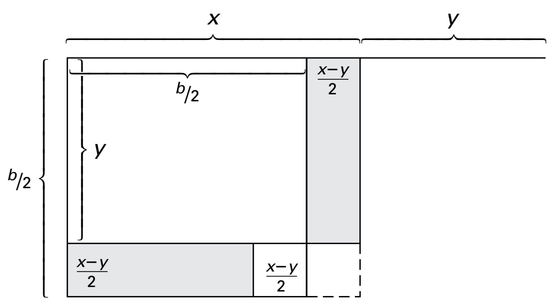


FIGURE 1.20

Geometric procedure for solving the system $x + y = b$, $xy = c$

Geometry is also at the base of the Babylonian solution of what we would consider a single quadratic equation. Several such problems are given on tablet BM 13901, including the following, where the translation shows the geometric flavor of the problem: "I summed the area and two-thirds of my square-side and it was 0;35. You put down 1, the projection. Two-thirds of 1, the projection, is 0;40. You combined its half, 0;20 and 0;20. You add 0;06,40 to 0;35 and 0;41,40 squares 0;50. You take away 0;20 that you combined from the middle of 0;50 and the square-side is 0;30."[16] In modern terms, the equation to be solved is $x^2 + (2/3)x = 7/12$. At first glance, it would appear that the statement of the problem is not a geometric one, since we are asked to add a multiple of a side to an area. But the word "projection" indicates that this two-thirds multiple of a side is to be considered as two-thirds of the rectangle of length 1 and unknown side x. For the solution, the scribe took half of 2/3 and squared it ("combine its half, 0;20 and 0;20"), then took the result 1/9 (or 0;06,40) and added it to 7/12 (0;35) to get 25/36 (0;41,40). The scribe then noted that 5/6 (0;50) is the square root of 25/36 ("0;41,40 squares 0;50"). He then subtracted the 1/3 from 5/6 to get the result 1/2 ("the square-side is 0;30"). The Babylonian rule exemplified by this problem is easily translated into a modern formula for solving $x^2 + bx = c$, namely,

$$x = \sqrt{(b/2)^2 + c} - b/2,$$

recognizable as a version of the quadratic formula. Figure 1.21 shows the geometric meaning of the procedure in the generic case, where we start with a square of side x adjoined by a rectangle of width x and length b. The procedure then amounts to cutting half of the rectangle off from one side of the square and moving it to the bottom. Adding a square of side $b/2$ "completes the square." It is then evident that the unknown length x is equal to the difference between the side of the new square and $b/2$, exactly as the formula implies.

For the analogous problem $x^2 - bx = c$, the Babylonian geometric procedure is equivalent to the formula $x = \sqrt{(b/2)^2 + c} + b/2$. This is illustrated by another problem from BM 13901, which we would translate as $x^2 - x = 870$: "I took away my square-side from inside the area and it was 14,30. You put down 1, the projection. You break off half of 1. You combine 0;30 and 0;30. You add 0;15 to 14,30. 14,30;15 squares 29;30. You add 0;30 which you combined to 29;30 so that the square-side is 30."[17]

One should, however, keep in mind that the "quadratic formula" did not mean the same thing to the Babylonian scribes as it means to us. First, the scribes gave different procedures

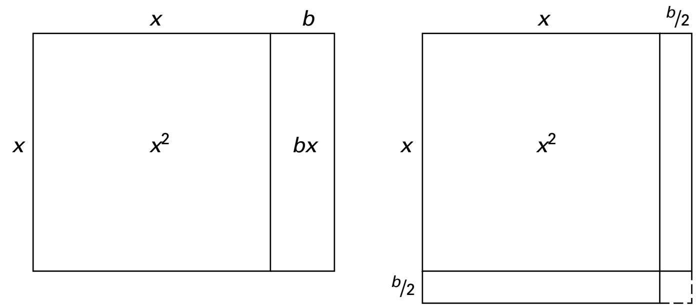


FIGURE 1.21

Geometric version of the quadratic formula for solving $x^2 + bx = c$

for solving the two types $x^2 + bx = c$ and $x^2 - bx = c$ because the two problems were different; they had different geometric meanings. To a modern mathematician, on the other hand, these problems are the same because the coefficient of x can be taken as positive or negative. Second, the modern quadratic formula in these two cases gives a positive and a negative solution to each equation. The negative solution, however, makes no geometrical sense and was completely ignored by the Babylonians.

In both of these quadratic equation problems, the coefficient of the x^2 term is 1. How did the Babylonians treat the quadratic equation $ax^2 \pm bx = c$ when $a \neq 1$? Again, there are problems on BM 13901 showing that the scribes scaled up the unknown to reduce the problem to the case $a = 1$. For example, problem 7 can be translated into the modern equation $11x^2 + 7x = 6\frac{1}{4}$. The scribe multiplied by 11 to turn the equation into a quadratic equation in $11x$: $(11x)^2 + 7 \cdot 11x = 68\frac{3}{4}$. He then solved

$$11x = \sqrt{\left(\frac{7}{2}\right)^2 + 68\frac{3}{4}} - \frac{7}{2} = \sqrt{81} - \frac{7}{2} = 9 - 3\frac{1}{2} = 5\frac{1}{2}.$$

To find x, the scribe would normally multiply by the reciprocal of 11, but in this case, he noted that the reciprocal of 11 "cannot be solved." Nevertheless, he realized, probably because the problem was manufactured to give a simple answer, that the unknown side x is equal to 1/2.

This idea of "scaling," combined with the geometrical coefficients discussed earlier, enabled the scribes to solve quadratic-type equations not directly involving squares. For example, consider the problem from TMS 20: The sum of the area and side of the convex square is 11/18. Find the side. We will translate this into the equation $A + s = 11/18$, where s is the quarter-circle arc forming one of the sides of the figure whose area is A. To solve this, the scribe used the coefficient 4/9 of the convex square as his scaling factor. Thus, he turned the equation into $(4/9)A + (4/9)s = 22/81$. But we know that the area A of the convex square is equal to $(4/9)s^2$. It follows that this equation can be rewritten as a quadratic equation for $(4/9)s$:

$$\left(\frac{4}{9}s\right)^2 + \frac{4}{9}s = \frac{22}{81}.$$

The scribe then solved this in the normal way to get $(4/9)s = 2/9$. He concluded by multiplying by the reciprocal 9/4 to find the answer $s = 1/2$.

Although the methods described above are the standard methods for solving quadratic equations, the scribes occasionally used other methods in particular situations. For example, in problem 23 of BM 13901, we are told that the sum of four sides and the (square) surface is 25/36. Although this problem is of the type $x^2 + bx = c$, in this case the b is four, the number of sides of the square, which is more "natural" than the coefficients we saw earlier. Modern scholars believe that this problem is an example of an original problem coming directly from the surveyors, a problem that then turns up in much later manifestations of this early tradition both in Islamic mathematics and in medieval European mathematics. The scribe's method here depends directly on the "four." In the first step of the solution, he took $1/4$ of the 25/36 to get $25/144$. To this he added 1, giving $169/144$. The square root of this value is $13/12$. Subtracting the 1 gives $1/12$. Thus, the length of the side is twice that value, namely, $1/6$. This new procedure is best illustrated by another diagram (Fig. 1.22). What the scribe intended is that the four "sides" are really projections of the actual sides of the square into rectangles of length 1. Taking $1/4$ of the entire sum means that we are only considering the shaded gnomon, which is one-fourth of the original figure. When we add a square of side 1 to that figure, we get a square whose side we can then find. Subtracting the 1 from the side then gives us half of the original side of the square.

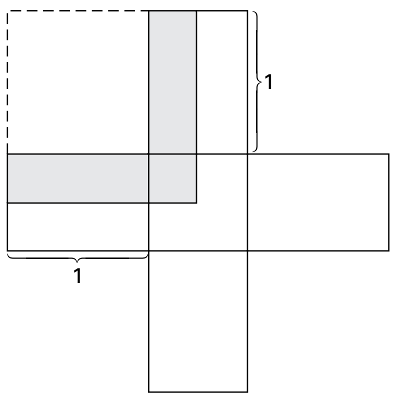


FIGURE 1.22
The sum of four sides and the square surface

Other problems on BM 13901 deal with various situations involving squares and sides, with each of the solution procedures having a geometric interpretation. As a final example, we consider the problem $x^2 + y^2 = 13/36$, $x - y = 1/6$. The solution to this system, which we generalize into the system $x^2 + y^2 = c$, $x - y = b$, was found by a procedure describable by the modern formula

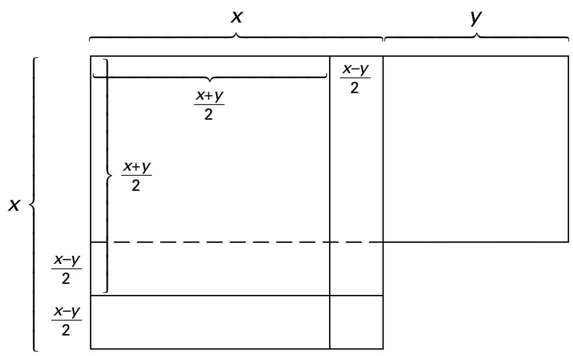
$$x = \sqrt{\frac{c}{2} - \left(\frac{b}{2}\right)^2} + \frac{b}{2} \qquad y = \sqrt{\frac{c}{2} - \left(\frac{b}{2}\right)^2} - \frac{b}{2}.$$

It appears that the Babylonians developed the solution by using the geometric idea expressed in Figure 1.23. This figure shows that

$$x^2 + y^2 = 2\left(\frac{x+y}{2}\right)^2 + 2\left(\frac{x-y}{2}\right)^2.$$

FIGURE 1.23

Geometric procedure for solving the system $x - y = b$, $x^2 + y^2 = c$



It follows that

$$c = 2\left(\frac{x+y}{2}\right)^2 + 2\left(\frac{b}{2}\right)^2$$

and therefore that

$$\frac{x+y}{2} = \sqrt{\frac{c}{2} - \left(\frac{b}{2}\right)^2}.$$

Because

$$x = \frac{x+y}{2} + \frac{x-y}{2} \quad \text{and} \quad y = \frac{x+y}{2} - \frac{x-y}{2},$$

the result follows.

1.3 CONCLUSION

The extant papyri and tablets containing Egyptian and Babylonian mathematics were generally teaching documents, used to transmit knowledge from one scribe to another. Their function was to provide trainee scribes with a set of example-types, problems whose solutions could be applied in other situations. Learning mathematics for these trainees was learning how to select and perhaps modify an appropriate algorithm, and then mastering the arithmetic techniques necessary to carry out the algorithm to solve a new problem. The reasoning behind the algorithms was evidently transmitted orally, so that mathematicians today are forced to speculate as to the origins.

We note that although the long lists of quadratic problems on some of the Babylonian tablets were given as "real-world" problems, the problems are in fact just as contrived as the ones found in most current algebra texts. That the authors knew they were contrived is shown by the fact that, typically, all problems of a given set have the same answer. But since often the problems grew in complexity, it appears that the tablets were used to develop techniques of solution. One can speculate, therefore, that the study of mathematical problem solving, especially problems involving quadratic equations, was a method for training the minds of future leaders of the country. In other words, it was not really that important to solve quadratic equations—there were few real situations that required them. What was important was for the students to develop skills in solving problems in general, skills that could be used in dealing with the everyday problems that a nation's leaders need to solve. These skills included not only following well-established procedures—algorithms—but also knowing how and when to modify the methods and how to reduce more complicated problems to ones already solved. Today's students are often told that mathematics is studied to "train the mind." It seems that teachers have been telling their students the same thing for the past 4000 years.

EXERCISES

1. Represent 375 and 4856 in Egyptian hieroglyphics and Babylonian cuneiform.
2. Use Egyptian techniques to multiply 34 by 18 and to divide 93 by 5.
3. Use Egyptian techniques to multiply $\overline{2}\ \overline{14}$ by $1\ \overline{2}\ \overline{4}$. (This is problem 9 of the *Rhind Mathematical Papyrus*.)
4. Use Egyptian techniques to multiply $\overline{28}$ by $1\ \overline{2}\ \overline{4}$. (This is problem 14 of the *Rhind Mathematical Papyrus*.)
5. Show that the solution to the problem of dividing 7 loaves among 10 men is that each man gets $\overline{\overline{3}}\ \overline{30}$. (This is problem 4 of the *Rhind Mathematical Papyrus*.)
6. Use Egyptian techniques to divide 100 by $7\ \overline{2}\ \overline{4}\ \overline{8}$. Show that the answer is $12\ \overline{\overline{3}}\ \overline{42}\ \overline{126}$. (This is problem 70 of the *Rhind Mathematical Papyrus*.)
7. Multiply $7\ \overline{2}\ \overline{4}\ \overline{8}$ by $12\ \overline{\overline{3}}$ using the Egyptian multiplication technique. Note that it is necessary to multiply each term of the multiplicand by $\overline{\overline{3}}$ separately.
8. A part of the *Rhind Mathematical Papyrus* table of division by 2 follows: $2 \div 11 = \overline{6}\ \overline{66}$, $2 \div 13 = \overline{8}\ \overline{52}\ \overline{104}$, $2 \div 23 = \overline{12}\ \overline{276}$. The calculation of $2 \div 13$ is given as follows:

1	13
$\overline{2}$	$6\ \overline{2}$
$\overline{4}$	$3\ \overline{4}$
$\overline{8}$	$1\ \overline{2}\ \overline{8}$
$\overline{52}$	$\overline{4}$
$\overline{104}$	$\overline{8}$
$\overline{8}\ \overline{52}\ \overline{104}$	$1\ \overline{2}\ \overline{4}\ \overline{8}\ \overline{8}$
	2

Perform similar calculations for the divisions of 2 by 11 and 23 to check the results.

9. Solve by the method of false position: A quantity and its 1/7 added together become 19. What is the quantity? (problem 24 of the *Rhind Mathematical Papyrus*)
10. Solve by the method of false position: A quantity and its 2/3 are added together and from the sum 1/3 of the sum is subtracted, and 10 remains. What is the quantity? (problem 28 of the *Rhind Mathematical Papyrus*)
11. A quantity, its 1/3, and its 1/4, added together, become 2. What is the quantity? (problem 32 of the *Rhind Mathematical Papyrus*)
12. Calculate a quantity such that if it is taken two times along with the quantity itself, the sum comes to 9. (problem 25 of the *Moscow Mathematical Papyrus*)
13. Problem 72 of the *Rhind Mathematical Papyrus* reads "100 loaves of *pesu* 10 are exchanged for loaves of *pesu* 45. How many of these loaves are there?" The solution is given as, "Find the excess of 45 over 10. It is 35. Divide this 35 by 10. You get $3\ \overline{2}$. Multiply $3\ \overline{2}$ by 100. Result: 350. Add 100 to this 350. You get 450. Say then that the exchange is 100 loaves of *pesu* 10 for 450 loaves of *pesu* 45."[18] Translate this solution into modern terminology. How does this solution demonstrate proportionality?
14. Solve problem 11 of the *Moscow Mathematical Papyrus*: The work of a man in logs; the amount of his work is 100 logs of 5 handbreadths diameter; but he has brought them in logs of 4 handbreadths diameter. How many logs of 4 handbreadths diameter are there?
15. Various conjectures have been made for the derivation of the Egyptian formula $A = (\frac{8}{9}d)^2$ for the area A of a circle of diameter d. One of these uses circular counters, known to have been used in ancient Egypt. Show by experiment using pennies, for example, whose diameter can be taken as 1, that a circle of diameter 9 can essentially be filled by 64 circles of diameter 1. (Begin with one penny in the center; surround it with a circle of six pennies, and so on.) Use the obvious fact that 64 circles of diameter 1 also fill a square

of side 8 to show how the Egyptians may have derived their formula.[19]

16. Some scholars have conjectured that the area calculated in problem 10 of the *Moscow Mathematical Papyrus* is that of a semicylinder rather than a hemisphere. Show that the calculation in that problem does give the correct surface area of a semicylinder of diameter and height both equal to $4\frac{1}{2}$.
17. Convert the fractions 7/5, 13/15, 11/24, and 33/50 to sexagesimal notation.
18. Convert the sexagesimal fractions 0;22,30, 0;08,06, 0;04,10, and 0;05,33,20 to ordinary fractions in lowest terms.
19. Find the reciprocals in base 60 of 18, 32, 54, and 64 (=1,04). (Do not worry about initial zeros, since the product of a number with its reciprocal can be any power of 60.) What is the condition on the integer n that ensures it is a regular sexagesimal, that is, that its reciprocal is a finite sexagesimal fraction?
20. In the Babylonian system, multiply 25 by 1,04 and 18 by 1,21. Divide 50 by 18 and 1,21 by 32 (using reciprocals). Use our standard multiplication algorithm modified for base 60.
21. Show that the area of the Babylonian "barge" is given by $A = (2/9)a^2$, where a is the length of the arc (one-quarter of the circumference). Also show that the length of the long transversal of the barge is $(17/18)a$ and the length of the short transversal is $(7/18)a$. (Use the Babylonian values of $C^2/12$ for the area of a circle and 17/12 for $\sqrt{2}$.)
22. Show that the area of the Babylonian "bull's-eye" is given by $A = (9/32)a^2$, where a is the length of the arc (one-third of the circumference). Also show that the length of the long transversal of the bull's-eye is $(7/8)a$, whereas the length of the short transversal is $(1/2)a$. (Use the Babylonian values of $C^2/12$ for the area of a circle and 7/4 for $\sqrt{3}$.)
23. For the concave square, the coefficient of the diagonal (the line from one vertex to the opposite vertex) is given as $1;20 (= 1\frac{1}{3})$, while the coefficient of the tranversal (the line from the midpoint of one arc to the midpoint of the opposite arc) is given as $0;33,20 (= 5/9)$. Show that both of these values are correct, given the normal Babylonian approximations.
24. Convert the Babylonian approximation 1;24,51,10 to $\sqrt{2}$ to decimals and determine the accuracy of the approximation.
25. Use the assumed Babylonian square root algorithm of the text to show that $\sqrt{3} \approx 1;45$ by beginning with the value 2. Find a three-sexagesimal-place approximation to the reciprocal of 1;45 and use it to calculate a three-sexagesimal-place approximation to $\sqrt{3}$.
26. Show that taking $v + u = 1;48\ (= 1\frac{4}{5})$ leads to line 15 of Plimpton 322 and that taking $v + u = 2;05\ (= 2\frac{1}{12})$ leads to line 9. Find the values for $v + u$ that lead to lines 6 and 13 of that tablet.
27. The scribe of Plimpton 322 did not use the value $v + u = 2;18,14,24$, with its associated reciprocal $v - u = 0;26,02,30$, in his work on the tablet. Find the smallest Pythagorean triple associated with those values.
28. Solve the problem from the Old Babylonian tablet BM 13901: The sum of the areas of two squares is 1525. The side of the second square is 2/3 that of the first plus 5. Find the sides of each square.
29. Solve the Babylonian problem taken from a tablet found at Susa: Let the width of a rectangle measure a quarter less than the length. Let 40 be the length of the diagonal. What are the length and width? Use false position, beginning with the assumption that 1 (or 60) is the length of the rectangle.
30. Solve the following problem from VAT 8391: One of two fields yields 2/3 *sila* per *sar*, the second yields 1/2 *sila* per *sar*. The sum of the yields of the two fields is 1100 *sila*; the difference of the areas of the two fields is 600 *sar*. How large is each field?
31. Solve the following problem from YBC 4652: I found a stone, but did not weigh it; after I subtracted one-seventh and then one-thirteenth [of the difference], it weighed 1 *mina* [= 60 *gin*]. What was the original weight of the stone?
32. Solve the following problem from YBC 4652: I found a stone, but did not weigh it; after I subtracted one-seventh, added one-eleventh [of the difference], and then subtracted one-thirteenth [of the previous total], it weighed 1 *mina* [= 60 *gin*]. What was the stone's weight?
33. Give a geometric argument to justify the Babylonian "quadratic formula" that solves the equation $x^2 - ax = b$.
34. Solve the following problem from tablet YBC 6967: A number exceeds its reciprocal by 7. Find the number and the reciprocal. (In this case, that two numbers are "reciprocals" means that their product is 60.)
35. Solve the following Babylonian problem about a concave square: The sum of the area, the arc, and the diagonal is $1;16,40 (= 1\frac{5}{18})$. Find the length of the arc. (Recall that the coefficient of the area is 4/9 and the coefficient of the diagonal is 1 1/3—see Exercise 23.)
36. Solve the following problem from BM 13901: I added one-third of the square-side to two-thirds of the area of the square, and the result was 0;20 (= 1/3). Find the square-side.

37. Solve the following Babylonian problem from tablet IM 55357: Given the right triangle ABC with sides 0;45 and 1 and hypotenuse 1;15, as in Figure 1.24, suppose AD is perpendicular to BC, DE is perpendicular to AC, and EF is perpendicular to BC. Suppose further that the area of triangle ABD is 0;08,06, that of triangle ADE is 0;05,11,02,24, that of triangle DEF is 0;03,19,03,56,09,36, and that of EFC is 0;05,53,53,39,50,24. What are the lengths of AD, DE, EF, BD, DF, and FC?

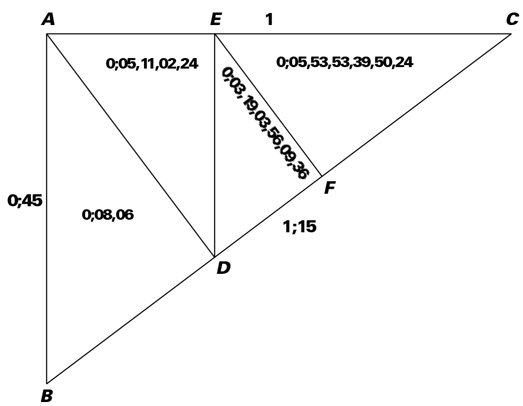


FIGURE 1.24

Tablet IM 55357 with a problem on triangles

38. Given a circle of circumference 60 and a chord of length 12, what is the perpendicular distance from the chord to the circumference? (This problem is from tablet BM 85194.)

39. Solve the following problem from tablet AO 8862: Length and width. I combined length and width and then I built an area. I turned around. I added half of the length and a third of the width to the middle of my area so that it was 15. I returned. I summed the length and width and it was 7. What are the length and width?

40. Construct two or three real-life division problems where giving the answer using just unit fractions, rather than other common fractions, makes sense.

41. Devise a lesson to teach ideas of proportionality by using the Egyptian method of false position.

42. Devise a lesson on place value using the Babylonian system and, in particular, using the multiplication table by 9 given in the text.

43. Devise a lesson teaching the quadratic formula using geometric arguments similar to the (assumed) Babylonian ones.

REFERENCES AND NOTES

The best source on Egyptian mathematics is Richard J. Gillings, *Mathematics in the Time of the Pharaohs* (Cambridge: MIT Press, 1972). See also Gillings, "The Mathematics of Ancient Egypt," *Dictionary of Scientific Biography* (New York: Scribners, 1978), vol. 15, 681–705. A more recent survey is James Ritter, "Egyptian Mathematics," in H. Selin, ed., *Mathematics across Cultures: The History of Non-Western Mathematics* (Dordrecht: Kluwer Academic Publishers, 2000), pp. 115–136. Finally, there is a new book in preparation on Egyptian mathematics by Annette Imhausen, soon to be published by Princeton University Press, which will probably supersede some of the earlier works. The *Rhind Mathematical Papyrus* is available in an edition by Arnold B. Chace (Reston, VA: National Council of Teachers of Mathematics, 1967). This work is an abridgement of the original publication by the Mathematical Association of America in 1927 and 1929. Another English version of that papyrus as well as the *Moscow Mathematical Papyrus* and other Egyptian mathematical fragments are available in Marshall Clagett, *Ancient Egyptian Science: A Source Book. Volume Three: Ancient Egyptian Mathematics* (Philadelphia: American Philosophical Society, 1999). New translations of some of these sources as well as a substantial number of Mesopotamian sources are in Victor Katz, ed., *The Mathematics of Egypt, Mesopotamia, China, India, and Islam: A Sourcebook* (Princeton: Princeton University Press, 2007).

The standard accounts of Babylonian mathematics are Otto Neugebauer, *The Exact Sciences in Antiquity* (Princeton: Princeton University Press, 1951; New York: Dover, 1969) and B. L. Van der Waerden, *Science Awakening I* (New York: Oxford University Press, 1961). More recent surveys are by Jens Høyrup, "Mathematics, Algebra, and Geometry," in *The Anchor Bible Dictionary*, David N. Freedman, ed. (New York: Doubleday, 1992), vol. IV, pp. 601–612, and by Jöran Friberg, "Mathematik," *Reallexikon der Assyriologie* 7 (1987–1990), pp. 531–585 (in English). Also, there will soon be a new book on Mesopotamian mathematics by Eleanor Robson, to be published by Princeton University Press, which will deal more with contextual issues than the earlier works cited here. It will undoubtedly be very much worth reading. Translations and analyses of the Babylonian tablets themselves are found principally in

Otto Neugebauer, *Mathematische Keilschrift-Texte* (New York: Springer, 1973, reprint of 1935 original), Otto Neugebauer and Abraham Sachs, *Mathematical Cuneiform Texts* (New Haven, CT: American Oriental Society, 1945), and Evert Bruins and M. Rutten, *Textes Mathématiques de Suse* (Paris: Paul Geuthner, 1961). Two technical works analyzing the Babylonian tablets are Jens Høyrup, *Lengths, Widths, Surfaces: A Portrait of Old Babylonian Algebra and Its Kin* (New York: Springer, 2002), and Eleanor Robson, *Mesopotamian Mathematics, 2100–1600 BC: Technical Constants in Bureaucracy and Education* (Oxford: Clarendon Press, 1999). More general surveys of Mesopotamian mathematics in the context of Mesopotamian society include chapter 3 of Jens Høyrup, *In Measure, Number, and Weight: Studies in Mathematics and Culture* (Albany: State University of New York Press, 1994), Eleanor Robson, "Mesopotamian Mathematics: Some Historical Background," in Victor Katz, ed., *Using History to Teach Mathematics* (Washington: MAA, 2000), pp. 149–158, and Eleanor Robson, "The Uses of Mathematics in Ancient Iraq, 6000–600 BC," in Selin, *Mathematics across Cultures*, pp. 93–113. Finally, the early Babylonian tokens are discussed in great detail and with many illustrations in Denise Schmandt-Besserat, *Before Writing: From Counting to Cuneiform* (Austin: University of Texas Press, 1992).

A book that discusses the mathematics of Egypt and Babylonia, along with that of other ancient societies comparatively, and also deals with the questions of transmission and a possible single origin of mathematics is B. L. Van der Waerden, *Geometry and Algebra in Ancient Civilizations* (New York: Springer, 1983).

1. Chace, *Rhind Mathematical Papyrus*, p. 27.
2. For the use of the zero in Egyptian architecture, see Dieter Arnold, *Building in Egypt* (New York: Oxford University Press, 1991), p. 17, and George Reisner, *Mycerinus: The Temples of the Third Pyramid at Giza* (Cambridge: Harvard University Press, 1931), pp. 76–77. For the use of the zero in Egyptian accounting, see Alexander Scharff, "Ein Rechnungsbuch des Königlichen Hofes aus der 13. Dynastie," *Zeitschrift für Ägyptische Sprache und Altertumskunde* 57 (1922), 58–59.
3. See Clagett, *Ancient Egyptian Science*, vol. 3, pp. 255–260, and Katz, *Sourcebook*, p. 21.
4. Clagett, *Ancient Egyptian Science*, vol. 3, p. 224.
5. Chace, *Rhind Mathematical Papyrus*, p. 69.
6. Gillings, *Mathematics in the Time of the Pharaohs*, p. 173.
7. Katz, *Sourcebook*, p. 39.
8. Gillings, *Mathematics in the Time of the Pharaohs*, p. 139. For further analysis, see Hermann Engels, "Quadrature of the Circle in Ancient Egypt," *Historia Mathematica* 4 (1977), 137–140.
9. Gillings, *Mathematics in the Time of the Pharaohs*, p. 188. See also Katz, *Sourcebook*, pp. 33–34, for further discussion.
10. Clagett, *Ancient Egyptian Science*, vol. 3, p. 218. See also Katz, *Sourcebook*, pp. 31–33, for a detailed discussion of this problem and its interpretation.
11. Mesopotamian tablets are generally, but not always, named by the library where they now reside. Thus, in the tablets mentioned in this chapter, we find W (Iraq Museum, Baghdad, Uruk excavations), YBC (Yale Babylonian Collection), BM (British Museum), Plimpton (Plimpton Collection at Columbia University), VAT (Vorderasiatischer Museum, Berlin), AO (Departement des Antiquités Orientales, Louvre, Paris), IM (Iraq Museum, Baghdad), and TMS (which stands for the book where these were first published: E. M. Bruins and M. Rutter, *Textes Mathématique de Suse* (Paris: Paul Geuthner, 1961)).
12. Robson, *Mesopotamian Mathematics*, pp. 50–54. This book has extensive discussions regarding geometrical coefficients.
13. Ibid., pp. 118–122. Besides the discussion here, there are translations of the tablets in which this and related grain piles appear on pp. 219–231.
14. This analysis of Plimpton 322 is based on Eleanor Robson's analysis as found in "Words and Pictures: New Light on Plimpton 322," *American Mathematical Monthly* 109 (2002), 105–120. A more technical discussion of the issues involved is in Eleanor Robson, "Neither Sherlock Holmes nor Babylon: A Reassessment of Plimpton 322," *Historia Mathematica* 28 (2001), 167–206. Both of these papers criticize an earlier assessment of Plimpton 322: R. Creighton Buck, "Sherlock Holmes in Babylon," *American Mathematical Monthly* 87 (1980), 335–345.
15. Neugebauer and Sachs, *Mathematical Cuneiform Texts*, p. 101.
16. See Katz, *Sourcebook*, p. 104. Also, see Jens Høyrup, "The Old Babylonian Square Texts—BM 13901 and YBC 4714," in Jens Høyrup and Peter Damerow, eds., *Changing Views on Ancient Near Eastern Mathematics* (Berlin: Dietrich Reimer Verlag, 2001). Both books have a complete translation of BM 13901, along with analyses of the various problems.
17. Ibid.
18. Chace, *Rhind Mathematical Papyrus*, p. 57.
19. This suggestion comes from Paulus Gerdes, "Three Alternate Methods of Obtaining the Ancient Egyptian Formula for the Area of a Circle," *Historia Mathematica* 12 (1985), 261–267. Two other possibilities are also presented in that article.

CHAPTER 2

The Beginnings of Mathematics in Greece

Thales was the first to go to Egypt and bring back to Greece this study [geometry]; he himself discovered many propositions, and disclosed the underlying principles of many others to his successors, in some cases his method being more general, in others more empirical.

—Proclus's *Summary* (c. 450 CE) of Eudemus's *History* (c. 320 BCE)[1]

A report from a visit to Egypt with Plato by Simmias of Thebes in 379 BCE (from a dramatization by Plutarch of Chaeronea (first–second century CE)): "On our return from Egypt a party of Delians met us . . . and requested Plato, as a geometer, to solve a problem set them by the god in a strange oracle. The oracle was to this effect: the present troubles of the Delians and the rest of the Greeks would be at an end when they had doubled the altar at Delos. As they not only were unable to penetrate its meaning, but failed absurdly in constructing the altar . . . , they called on Plato for help in their difficulty. Plato . . . replied that the god was ridiculing the Greeks for their neglect of education, deriding, as it were, our ignorance and bidding us engage in no perfunctory study of geometry; for no ordinary or near-sighted intelligence, but one well versed in the subject, was required to find two mean proportionals, that being the only way in which a body cubical in shape can be doubled with a similar increment in all dimensions. This would be done for them by Eudoxus of Cnidus . . . ; they were not, however, to suppose that it was this the god desired, but rather that he was ordering the entire Greek nation to give up war and its miseries and cultivate the Muses, and by calming their passions through the practice of discussion and study of mathematics, so to live with one another that their relationships should be not injurious, but profitable."[2]

As the quotation and the (probably) fictional account indicate, a new attitude toward mathematics appeared in Greece sometime before the fourth century BCE. It was no longer sufficient merely to calculate numerical answers to problems. One now had to prove that the results were correct. To double a cube, that is, to find a new cube whose volume was twice that of the original one, is equivalent to determining the cube root of 2, and that was not a difficult problem numerically. The oracle, however, was not concerned with numerical calculation, but with geometric construction. That in turn depended on geometric proof by some logical argument, the earliest manifestation of such in Greece being attributed to Thales.

This change in the nature of mathematics, beginning around 600 BCE, was related to the great differences between the emerging Greek civilization and those of Egypt and Babylonia, from whom the Greeks learned. The physical nature of Greece with its many mountains and islands is such that large-scale agriculture was not possible. Perhaps because of this, Greece did not develop a central government. The basic political organization was the *polis,* or city-state. The governments of the city-state were of every possible variety but in general controlled populations of only a few thousand. Whether the governments were democratic or monarchical, they were not arbitrary. Each government was ruled by law and therefore encouraged its citizens to be able to argue and debate. It was perhaps out of this characteristic that there developed the necessity for proof in mathematics, that is, for argument aimed at convincing others of a particular truth.

Because virtually every city-state had access to the sea, there was constant trade, both in Greece itself and with other civilizations. As a result, the Greeks were exposed to many different peoples and, in fact, themselves settled in areas all around the eastern Mediterranean. In addition, a rising standard of living helped to attract able people from other parts of the world. Hence, the Greeks were able to study differing answers to fundamental questions about the world. They began to create their own answers. In many areas of thought, they learned not to accept what had been handed down from ancient times. Instead, they began to ask, and to try to answer, "Why?" Greek thinkers eventually came to the realization that the world around them was knowable, that they could discover its characteristics by rational inquiry. Hence, they were anxious to discover and expound theories in such fields as mathematics, physics, biology, medicine, and politics. And although Western civilization owes a great debt to Greek society in literature, art, and architecture, it is to Greek mathematics that we owe the idea of mathematical proof, an idea at the basis of modern mathematics and, by extension, at the foundation of our modern technological civilization.

This chapter discusses the Greek numerical system and then considers the contributions of the earliest Greek mathematicians beginning in the sixth century BCE. It then deals with the beginnings of the Greek approach to geometric problem solving and concludes with the work of Plato and Aristotle in the fourth century BCE on the nature of mathematics and the idea of logical reasoning.

2.1 THE EARLIEST GREEK MATHEMATICS

Unlike the situation with Egyptian and Babylonian mathematics, there are virtually no extant texts of Greek mathematics that were actually written in the first millennium BCE. What we have today are copies of copies of copies, where the actual written documents date from not much earlier than 1000 CE. And even then, the earliest complete texts (of which these are copies) are not from earlier than about 300 BCE. So to tell the story of early Greek mathematics, we are forced to rely on works that were originally written much later than the actual occurrences. Thus, given that these works do not always agree with each other, there is a considerable amount of controversy about some of the early developments. We will try to present the story as coherently as possible, but will note many areas in which scholarly opinion varies.

2.1.1 Greek Numbers

From what fragments exist from ancient times, and even from some of the copies, we do know that the Greeks represented numbers in a ciphered system using their alphabet, from as far back as the sixth century BCE. The representation was as shown in Table 2.1, where the letters ϛ (digamma) for 6, ϙ (koppa) for 90, and ϡ (sampi) for 900 are letters that by this time were no longer in use. Hence, 754 was written $\psi\nu\delta$ and 293 was written σϙγ. To represent thousands, a mark was made to the left of the letters α through θ; for example, $'\theta$ represented 9000. Larger numbers still were written using the letter M to represent myriads (10,000), with the number of myriads written above: $\overset{\delta}{M} = 40{,}000$, $\overset{'\zeta\rho o\epsilon}{M} = 71{,}750{,}000$.

TABLE 2.1 ***Representation of a number system used by the Greeks as early as the sixth century BCE.***

Letter	Value	Letter	Value	Letter	Value
α	1	ι	10	ρ	100
β	2	κ	20	σ	200
γ	3	λ	30	τ	300
δ	4	μ	40	υ	400
ϵ	5	ν	50	ϕ	500
ϛ	6	ξ	60	χ	600
ζ	7	o	70	ψ	700
η	8	π	80	ω	800
θ	9	ϙ	90	ϡ	900

Among the earliest extant inscriptions in this alphabetic cipher were numbers inscribed on the walls of the tunnel on the island of Samos constructed by Eupalinus around 550 BCE to bring water from a spring outside the capital city through a mountain to a point inside the city walls. Modern archaeological excavations of the tunnel have revealed that it was dug by two teams that met in the middle (Fig. 2.1). There are no records as to how the construction crews managed to keep digging in the correct direction, but there have been many theories as to how this was done. The latest archaeological evidence leads to the conclusion that the builders used the simplest possible mathematical techniques, such as lining up flags to make sure that the diggers kept digging in the right direction. And evidently the numbers on the walls, 10, 20, 30, . . . , 200 (from the south entrance) and 10, 20, 30, . . . , 300 (from the north entrance) were written to keep tabs on the distances dug. Although most of the tunnel is

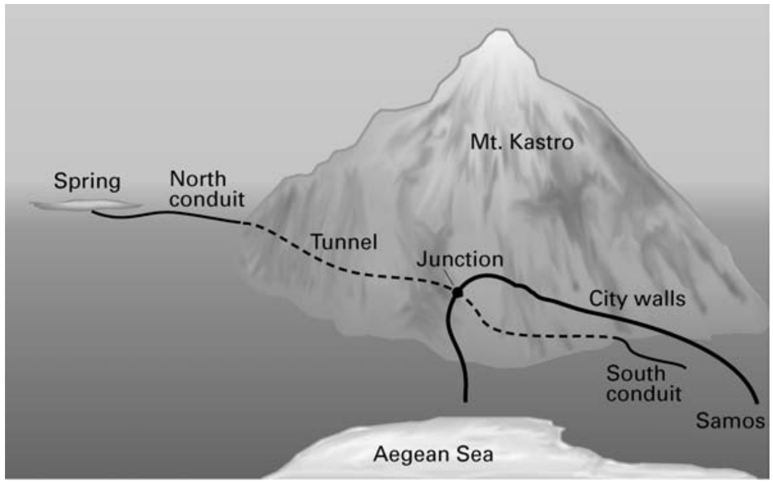


FIGURE 2.1
Water tunnel on the island of Samos

straight, there is one clear jog in the tunnel, probably necessitated by difficult soil conditions. Somehow, Eupalinus managed to figure out at that point how to get the digging back to the correct direction.

The numbers in the Eupalinus tunnel are integers. But Greek merchants and accountants, for example, needed fractions as well. Generally, in this early period, the Greeks used the Egyptian system of "parts." There was a special symbol $\angle$, which represented a half; $\acute{\beta}$ represented two-thirds. For the rest, the system was standard: $\acute{\gamma}$ represented one-third, $\acute{\delta}$ one-fourth, and so on. More complicated fractions than simple parts are expressed as the sum of an integer and different simple parts. For example, the fraction we represent as 12/17 might be represented as $\angle\acute{\iota}\acute{\beta}\acute{\iota}\acute{\zeta}\acute{\lambda}\acute{\delta}\acute{\nu}\acute{\alpha}\acute{\xi}\acute{\eta}$, which in modern notation would be $\frac{1}{2}+\frac{1}{12}+\frac{1}{17}+\frac{1}{34}+\frac{1}{51}+\frac{1}{68}$. We do not know if there was any systematic method for figuring out which unit fractions to use, for there are many possible ways to represent 12/17, or as the Greeks would say, the "seventeenth part of twelve." In addition, there is clearly the possibility of confusion between the representations of, for example, $\frac{1}{20}+\frac{1}{5}$ and $\frac{1}{25}$. But all those who needed to calculate evidently had methods of determining how they would use this system and how to avoid confusion.[3]

Fortunately for us, most of the early Greek mathematics we will discuss involves little calculation. As Aristotle wrote in his *Metaphysics*,

> At first, he who invented any art whatever that went beyond the common perceptions of man was naturally admired by men, not only because there was something useful in the inventions, but because he was thought wise and superior to the rest. But as more arts were invented, and some were directed to the necessities of life, others to recreation, the inventors of the latter were naturally always regarded as wiser than the inventors of the former, because their branches of knowledge did not aim at utility. Hence when all such inventions were already established, the sciences which do not aim at giving pleasure or at the necessities of life were discovered, and first in the places where men first began to have leisure. This is why the mathematical arts were founded in Egypt; for there the priestly caste was allowed to be at leisure.[4]

Although Aristotle referred only to Egypt, he certainly believed that in Greece as well mathematics was the province of a leisured class, people who did not deal with such mundane matters as measurement or accountancy problems. Thus, in Greece as in Egypt and Mesopotamia, mathematics of the type we will discuss in this chapter and the next was the province of a very limited group of people, virtually all of whom were part of the ruling groups. As we will see, this theoretical mathematics was to be a central part of the education of the rulers of the state.

2.1.2 Thales

FIGURE 2.2
Thales on a Greek stamp

The most complete reference to the earliest Greek mathematics is in the commentary to Book I of Euclid's *Elements* written in the fifth century CE by Proclus, some 800 to 1000 years after the fact. This account of the early history of Greek mathematics is generally thought to be a summary of a formal history written by Eudemus of Rhodes in about 320 BCE, the original of which is lost. In any case, the earliest Greek mathematician mentioned is Thales (c. 624–547 BCE), from Miletus in Asia Minor (Fig. 2.2). There are many stories recorded about him, most written down several hundred years after his death. These include his prediction of a solar eclipse in 585 BCE and his application of the angle-side-angle criterion of triangle congruence to the problem of measuring the distance to a ship at sea. He is said to have impressed Egyptian officials by determining the height of a pyramid by comparing the length of its shadow to that of the length of the shadow of a stick of known height. Thales is also credited with discovering the theorems that the base angles of an isosceles triangle are equal and that vertical angles are equal and with proving that the diameter of a circle divides the circle into two equal parts. Although exactly how Thales "proved" any of these results is not known, it does seem clear that he advanced some logical arguments.

Aristotle related the story that Thales was once reproved for wasting his time on idle pursuits. Therefore, noticing from certain signs that a bumper crop of olives was likely in a particular year, he quietly cornered the market on oil presses. When the large crop in fact was harvested, the olive growers all had to come to him for presses. He thus demonstrated that a philosopher or a mathematician could in fact make money if he thought it worthwhile. Whether this or any of the other stories are literally true is not known. In any case, the Greeks of the fourth century BCE and later credited Thales with beginning the Greek mathematical tradition. In fact, he is generally credited with beginning the entire Greek scientific enterprise, including recognizing that material phenomena are governed by discoverable laws.

2.1.3 Pythagoras and His School

FIGURE 2.3
Pythagoras on a Greek coin

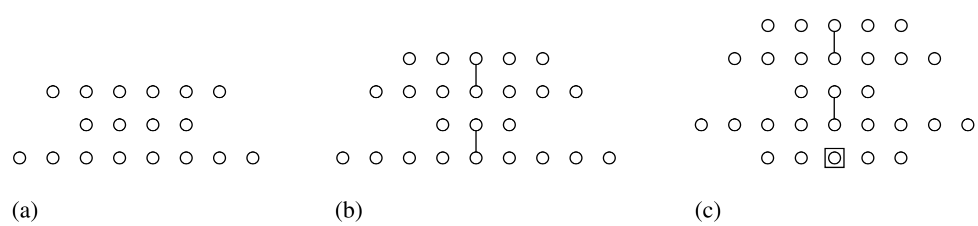
There are also extensive but unreliable stories about Pythagoras (c. 572–497 BCE), including that he spent much time not only in Egypt, where Thales was said to have visited, but also in Babylonia (Fig. 2.3). Around 530 BCE, after having been forced to leave his native Samos, he settled in Crotona, a Greek town in southern Italy. There he gathered around him a group of disciples, later known as the Pythagoreans, in what was considered both a religious order and a philosophical school. From the surviving biographies, all written centuries after his death, we can infer that Pythagoras was probably more of a mystic than a rational thinker, but one who commanded great respect from his followers. Since there are no extant works ascribed to Pythagoras or the Pythagoreans, the mathematical doctrines of his school can only be surmised from the works of later writers, including the "neo-Pythagoreans."

One important such mathematical doctrine was that "number was the substance of all things," that numbers, that is, positive integers, formed the basic organizing principle of the universe. What the Pythagoreans meant by this was not only that all known objects have a number, or can be ordered and counted, but also that numbers are at the basis of all physical phenomena. For example, a constellation in the heavens could be characterized by both the number of stars that compose it and its geometrical form, which itself could be thought of as represented by a number. The motions of the planets could be expressed in terms of ratios of numbers. Musical harmonies depend on numerical ratios: two plucked strings with ratio of length $2:1$ give an octave, with ratio $3:2$ give a fifth, and with ratio $4:3$ give a fourth. Out of these intervals an entire musical scale can be created. Finally, the fact that triangles whose sides are in the ratio of $3:4:5$ are right-angled established a connection of number with angle. Given the Pythagoreans' interest in number as a fundamental principle of the cosmos, it is only natural that they studied the properties of positive integers, what we would call the elements of the theory of numbers.

The starting point of this theory was the dichotomy between the odd and the even. The Pythagoreans probably represented numbers by dots or, more concretely, by pebbles. Hence, an even number would be represented by a row of pebbles that could be divided into two equal parts. An odd number could not be so divided because there would always be a single pebble left over. It was easy enough using pebbles to verify some simple theorems. For example, the sum of any collection of even numbers is even, while the sum of an even collection of odd numbers is even and that of an odd collection is odd (Fig. 2.4).

FIGURE 2.4

(a) The sum of even numbers is even. (b) An even sum of odd numbers is even. (c) An odd sum of odd numbers is odd.



Among other simple corollaries of the basic results above were the theorems that the square of an even number is even, while the square of an odd number is odd. Squares themselves could also be represented using dots, providing simple examples of "figurate" numbers. If one represents a given square in this way, for example, the square of 4, it is easy to see that the next higher square can be formed by adding a row of dots around two sides of the original figure. There are $2 \cdot 4 + 1 = 9$ of these additional dots. The Pythagoreans generalized this observation to show that one can form squares by adding the successive odd numbers to 1. For example, $1 + 3 = 2^2$, $1 + 3 + 5 = 3^2$, and $1 + 3 + 5 + 7 = 4^2$. The added odd numbers were in the L shape generally called a gnomon (Fig. 2.5). Other examples of figurate numbers include the triangular numbers, also shown in Figure 2.5, produced by successive additions of the natural numbers themselves. Similarly, oblong numbers, numbers of the form $n(n + 1)$, are produced by beginning with 2 and adding the successive even numbers (Fig. 2.6). The first

FIGURE 2.5

Square and triangular numbers

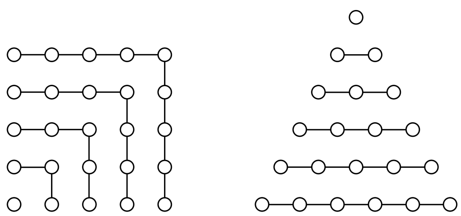


FIGURE 2.6

Oblong numbers

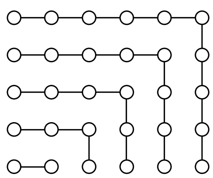
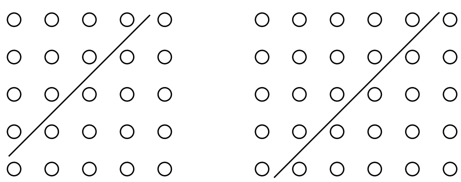


FIGURE 2.7

Two theorems on triangular numbers



four of these are 2, 6, 12, and 20, that is, 1×2, 2×3, 3×4, and 4×5. Figure 2.7 provides easy demonstrations of the results that any oblong number is the double of a triangular number and that any square number is the sum of two consecutive triangular numbers.

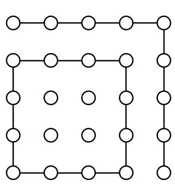


FIGURE 2.8

An odd square that is the difference of two squares

Another number theoretical problem of particular interest to the Pythagoreans was the construction of Pythagorean triples. There is evidence that they saw that for an odd number n, the triple $(n,\ \frac{n^2-1}{2},\ \frac{n^2+1}{2})$ is a Pythagorean triple, while if m is even, $(m,\ (\frac{m}{2})^2 - 1,\ (\frac{m}{2})^2 + 1)$ is such a triple. An explanation of how the Pythagoreans may have demonstrated the first of these results from their dot configurations begins with the remark that any odd number is the difference of two consecutive squares. Hence, if the odd number is itself a square, then three square numbers have been found such that the sum of two equals the third (Fig. 2.8). To find the sides of these squares, the Pythagorean triple itself, note that the side of the gnomon is given since it is the square of an odd number. The side of the smaller square is found by subtracting 1 from the gnomon and halving the remainder. The side of the larger square is one more than that of the smaller. A similar proof can be given for the second result. Although there is no explicit testimony to additional results involving Pythagorean triples, it seems probable that the Pythagoreans considered the odd and even properties of these triples. For example, it is not difficult to prove that in a Pythagorean triple, if one of the terms is odd, then two of them must be odd and one even.

The geometric theorem out of which the study of Pythagorean triples grew, namely, that in any right triangle the square on the hypotenuse is equal to the sum of the squares on the legs, has long been attributed to Pythagoras himself, but there is no direct evidence of this. The theorem was known in other cultures long before Pythagoras lived. Nevertheless, it was the knowledge of this theorem by the fifth century BCE that led to the first discovery of what is today called an irrational number.

For the early Greeks, number always was connected with things counted. Because counting requires that the individual units must remain the same, the units themselves can never be divided or joined to other units. In particular, throughout formal Greek mathematics, a number meant a "multitude composed of units," that is, a counting number. Furthermore, since the unit 1 was not a multitude composed of units, it was not considered a number in the same sense as the other positive integers. Even Aristotle noted that two was the smallest "number."

Because the Pythagoreans considered number as the basis of the universe, everything could be counted, including lengths. In order to count a length, of course, one needed a measure. The Pythagoreans thus assumed that one could always find an appropriate measure. Once such a measure was found in a particular problem, it became the unit and thus could not be divided. In particular, the Pythagoreans assumed that one could find a measure by which both the side and diagonal of a square could be counted. In other words, there should exist a length such that the side and diagonal were integral multiples of it. Unfortunately, this turned out not to be true. The side and diagonal of a square are **incommensurable**; there is no common measure. Whatever unit of measure is chosen such that an exact number will fit the length of one of these lines, the other line will require some number plus a portion of the unit, and one cannot divide the unit. (In modern terms, this result is equivalent to the statement that the square root of two is irrational.) We do not know who discovered this result, but scholars believe that the discovery took place in approximately 430 BCE. And although it is frequently stated that this discovery precipitated a crisis in Greek mathematics, the only reliable evidence shows that the discovery simply opened up the possibility of some new mathematical theories. In fact, Aristotle wrote in his *Metaphysics*,

> For all men begin, as we said, by wondering that things are as they are, as they do about self-moving marionettes, or about the solstices or the incommensurability of the diagonal of a square with the side; for it seems wonderful to all who have not yet seen the reason, that there is a thing which cannot be measured even by the smallest unit. But we must end in the contrary and, according to the proverb, the better state, as is the case in these instances too when men learn the cause; for there is nothing which would surprise a geometer so much as if the diagonal turned out to be commensurable.[5]

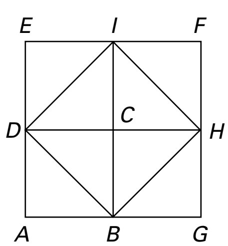


FIGURE 2.9

The incommensurability of the side and idagonal of a square (first possibility)

In other words, Aristotle seems to say that although the incommensurability is initially surprising, once one finds the reason—and clearly Greek thinkers did so—it then becomes very unsurprising.

So what is the "cause" of the incommensurability and how did a Greek thinker discover it? The only hint is in another work of Aristotle, who notes that if the side and diagonal are assumed commensurable, then one may deduce that odd numbers equal even numbers. One possibility as to the form of the discovery is the following: Assume that the side BD and diagonal DH in Figure 2.9 are commensurable, that is, that each is represented by the number of times it is measured by their common measure. It may be assumed that at least one of these numbers is odd, for otherwise there would be a larger common measure. Then the squares $DBHI$ and $AGFE$ on the side and diagonal, respectively, represent square numbers. The latter square is clearly double the former, so it represents an even square number. Therefore, its side $AG = DH$ also represents an even number and the square $AGFE$ is a multiple of four. Since $DBHI$ is half of $AGFE$, it must be a multiple of two; that is, it represents an even square. Hence, its side BD must also be even. But this contradicts the original assumption, that one of DH, BD, must be odd. Therefore, the two lines are incommensurable.

It must be realized that such a proof presupposes that by this time the notion of proof was ingrained into the Greek conception of mathematics. Although there is no evidence that the Greeks of the fifth century BCE possessed the entire mechanism of an axiomatic system and had explicitly recognized that certain statements need to be accepted without proof, they certainly had decided that some form of logical argument was necessary for determining the truth of a particular result. Furthermore, this entire notion of incommensurability represents a break from the Babylonian and Egyptian concepts of calculation with numbers. There is naturally no question that one can assign a numerical value to the length of the diagonal of a square of side one unit, as the Babylonians did, but the notion that no "exact" value can be found is first formally recognized in Greek mathematics.

Although the Greeks could not "measure" the diagonal of a square, that line, as a geometric object, was still significant. Plato, in his dialogue *Meno*, had Socrates question a slave boy about finding a square whose area is double that of square of side two feet. The boy first suggests that each side should be doubled. Socrates pointed out that this would give a square of area sixteen. The boy's second guess, that the new side should be three feet, is also evidently incorrect. So Socrates then led him to figure out that if one draws a diagonal of the original square and then constructs a square on that diagonal, the new square is exactly double the old one. But Socrates' proof of this is simply by a dissection argument (Fig. 2.10). There is no mention of the length of this diagonal at all.[6]

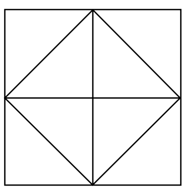


FIGURE 2.10

Dissection argument for determining the diagonal of a square

2.1.4 Squaring the Circle and Doubling the Cube

The idea of proof and the change from numerical calculation are further exemplified in the mid-fifth century attempts to solve two geometric problems, problems that were to occupy Greek mathematicians for centuries: the squaring of the circle (already attempted in Egypt) and the duplication of the cube (as noted in the oracle). The multitude of attacks on these particular problems and the slightly later one of trisecting an arbitrary angle serve to remind us that a central goal of Greek mathematics was geometrical problem solving, and that, to a large extent, the great body of theorems found in the major extant works of Greek mathematics served as logical underpinnings for these solutions. Interestingly, that these problems apparently could not be solved via the original tools of straightedge and compass was known to enough of the Greek public that Aristophanes could refer to "squaring the circle" as something absurd in his play *The Birds*, first performed in 414 BCE.

Hippocrates of Chios (mid-fifth century BCE) (no connection to the famous physician) was among the first to attack the cube and circle problems. As to the first of these, Hippocrates perhaps realized that the problem was analogous to the simpler problem of doubling a square of side a. That problem could be solved by constructing a mean proportional b between a and $2a$, a length b such that $a : b = b : 2a$, for then $b^2 = 2a^2$. From the fragmentary records of Hippocrates' work, it is evident that he was familiar with performing such constructions. In any case, ancient accounts record that Hippocrates was the first to come up with the idea of reducing the problem of doubling the cube of side a to the problem of finding two mean proportionals b, c, between a and $2a$. For if $a : b = b : c = c : 2a$, then

$$a^3 : b^3 = (a : b)^3 = (a : b)(b : c)(c : 2a) = a : 2a = 1 : 2$$

and $b^3 = 2a^3$. Hippocrates was not, however, able to construct the two mean proportionals using the geometric tools at his disposal. It was left to some of his successors to find this construction.

Hippocrates similarly made progress in the squaring of the circle, essentially by showing that certain lunes (figures bounded by arcs of two circles) could be "squared," that is, that their areas could be shown equal to certain regions bounded by straight lines. To do this, he first had to show that the areas of circles are to one another as the squares on their diameters, a fact evidently known to the Babylonian scribes. How he accomplished this is not known. In any case, he could now square the lune on a quadrant of a circle.

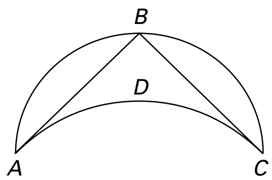


FIGURE 2.11

Hippocrates' lune on an isosceles right triangle

Suppose that semicircle ABC is circumscribed about the isosceles right triangle ABC and that around the base AC an arc ADC of a circle is drawn so that segment ADC is similar to segments AB and BC; that is, the arcs of each are the same fraction of a circle, in this case, one-quarter (Fig. 2.11). It follows from the result on areas of circles that similar segments are also to one another as the squares on their chords. Therefore, segment ADC is equal to the sum of segments AB and BC. If we add to each of these areas the part of the triangle outside arc ADC, it follows that the lune $ABCD$ is equal to the triangle ABC.

Although Hippocrates gave constructions for squaring other lunes or combinations of lunes, he was unable to actually square a circle. Nevertheless, it is apparent that his attempts on the squaring problem and the doubling problem were based on a large collection of geometric theorems, theorems that he organized into the first recorded book on the elements of geometry.

2.2 THE TIME OF PLATO

The time of Plato (429–347 BCE) (Fig. 2.12) saw significant efforts made toward solving the problems of doubling the cube and squaring the circle and toward dealing with incommensurability and its impact on the theory of proportion. These advances were achieved partly because Plato's Academy, founded in Athens around 385 BCE, drew together scholars from all over the Greek world. These scholars conducted seminars in mathematics and philosophy with small groups of advanced students and also conducted research in mathematics, among other fields. There is an unverifiable story, dating from some 700 years after the school's founding, that over the entrance to the Academy was inscribed the Greek phrase ΑΓΕΩΜΕΤΡΗΤΟΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ, meaning roughly, "Let no one ignorant of geometry enter here." A student "ignorant of geometry" would also be ignorant of logic and hence unable to understand philosophy.

FIGURE 2.12

Plato and Aristotle: A detail of Raphael's painting *The School of Athens*

The mathematical syllabus inaugurated by Plato for students at the Academy is described by him in his most famous work, *The Republic*, in which he discussed the education that should be received by the philosopher-kings, the ideal rulers of a state. The mathematical part of this education was to consist of five subjects: arithmetic (that is, the theory of numbers), plane geometry, solid geometry, astronomy, and harmonics (music). The leaders of the state are "to practice calculation, not like merchants or shopkeepers for purposes of buying and selling, but with a view to war and to help in the conversion of the soul itself from the world of becoming to truth and reality. . . . It will further our intentions if it is pursued for the sake of knowledge and not for commercial ends. . . . It has a great power of leading the mind upwards and forcing it to reason about pure numbers, refusing to discuss collections of material things which can be seen and touched."[7] In other words, arithmetic is to be studied for the training of the mind (and incidentally for its military usefulness). The arithmetic of which Plato writes includes not only the Pythagorean number theory already discussed but also additional material that is included in Books VII–IX of Euclid's *Elements* and will be considered later.

Again, a limited amount of plane geometry is necessary for practical purposes, particularly in war, when a general must be able to lay out a camp or extend army lines. But even though mathematicians talk of operations in plane geometry such as squaring or adding, the object of geometry, according to Plato, is not to *do* something but to gain knowledge, "knowledge, moreover, of what eternally exists, not of anything that comes to be this or that at some time and ceases to be."[8] So, as in arithmetic, the study of geometry—and for Plato this means theoretical, not practical, geometry—is for "drawing the soul towards truth." It is importanat to mention here that Plato distinguished carefully between, for example, the real geometric circles drawn by people and the essential or ideal circle, held in the mind, which is the true object of geometric study. In practice, one cannot draw a circle and its tangent with only one point in common, although this is the nature of the mathematical circle and the mathematical tangent.

The next subject of mathematical study should be solid geometry. Plato complained in the *Republic* that this subject has not been sufficiently investigated. This is because "no state thinks [it] worth encouraging" and because "students are not likely to make discoveries without a director, who is hard to find."[9] Nevertheless, Plato felt that new discoveries would be made in this field, and, in fact, much was done between the dramatic date of the dialogue (about 400 BCE) and the time of Euclid, some of which is included in Books XI–XIII of the *Elements*.

In any case, a decent knowledge of solid geometry was necessary for the next study, that of astronomy, or, as Plato puts it, "solid bodies in circular motion." Again, in this field Plato distinguished between the stars as material objects with motions showing accidental irregularities and variations and the ideal abstract relations of their paths and velocities expressed in numbers and perfect figures such as the circle. It is this mathematical study of ideal bodies that is the true aim of astronomical study. Thus, this study should take place by means of problems and without attempting to actually follow every movement in the heavens.

Similarly, a distinction is made in the final subject, of harmonics, between material sounds and their abstraction. The Pythagoreans had discovered the harmonies that occur when strings are plucked together with lengths in the ratios of certain small positive integers. But in encouraging his philosopher-kings in the study of harmonics, Plato meant for them to go beyond the actual musical study, using real strings and real sounds, to the abstract level of "inquiring which numbers are inherently consonant and which are not, and for what reasons."[10] That is, they should study the mathematics of harmony, just as they should study the mathematics of astronomy, and should not be overly concerned with real stringed instruments or real stars. It turns out that a principal part of the mathematics necessary in both studies is the theory of ratio and proportion, the subject matter of Euclid's *Elements*, Book V.

Although it is not known whether the entire syllabus discussed by Plato was in fact taught at the Academy, it is certain that Plato brought in the best mathematicians of his day to teach and do research, including Theaetetus (c. 417–369 BCE) and Eudoxus (c. 408–355 BCE), who we will discuss later. The most famous person associated with the Academy, however, was Aristotle.

2.3 ARISTOTLE

Aristotle (384–322 BCE) (Fig. 2.13) studied at Plato's Academy in Athens from the time he was 18 until Plato's death in 347. Shortly thereafter, he was invited to the court of Philip II of Macedon to undertake the education of Philip's son Alexander, who soon after his own accession to the throne in 335 began his successful conquest of the Mediterranean world (Fig. 2.14). Meanwhile, Aristotle returned to Athens where he founded his own school, the Lyceum, and spent the rest of his days writing, lecturing, and holding discussions with his advanced students. Although Aristotle wrote on many subjects, including politics, ethics, epistemology, physics, and biology, his strongest influence as far as mathematics was concerned was in the area of logic.

FIGURE 2.13

Bust of Aristotle

FIGURE 2.14

Painting of Alexander on horseback

2.3.1 Logic

Although there is only fragmentary evidence of logical argument in mathematical works before the time of Euclid, some appearing in the work of Hippocrates already mentioned, it is apparent that from at least the sixth century BCE, the Greeks were developing the notions of logical reasoning. The active political life of the city-states encouraged the development of argumentation and techniques of persuasion. And there are many examples from philosophical works, especially those of Parmenides (late sixth century BCE) and his disciple Zeno of Elea (fifth century BCE), that demonstrate various detailed techniques of argument. In particular, there are examples of such techniques as *reductio ad absurdum*, in which one assumes that a proposition to be proved is false and then derives a contradiction, and *modus tollens*, in which one shows first that if A is true, then B follows, shows next that B is not true, and concludes finally that A is not true. It was Aristotle, however, who took the ideas developed over the centuries and first codified the principles of logical argument.

Aristotle believed that logical arguments should be built out of **syllogisms**, where "a syllogism is discourse in which, certain things being stated, something other than what is stated follows of necessity from their being so."[11] In other words, a syllogism consists of certain statements that are taken as true and certain other statements that are then necessarily true. For example, the argument "if all monkeys are primates, and all primates are mammals, then it follows that all monkeys are mammals," exemplifies one type of syllogism, whereas the argument "if all Catholics are Christians and no Christians are Moslem, then it follows that no Catholic is Moslem," exemplifies a second type.

After clarifying the principles of dealing with syllogisms, Aristotle noted that syllogistic reasoning enables one to use "old knowledge" to impart new. If one accepts the premises of a syllogism as true, then one must also accept the conclusion. One cannot, however, obtain every piece of knowledge as the conclusion of a syllogism. One has to begin somewhere with truths that are accepted without argument. Aristotle distinguished between the basic truths that are peculiar to each particular science and the ones that are common to all. The former are often called **postulates**, while the latter are known as **axioms**. As an example of a common truth, he gave the axiom "take equals from equals and equals remain." His examples of peculiar truths for geometry are "the definitions of line and straight." By these he presumably meant that one postulates the existence of straight lines. Only for the most basic ideas did Aristotle permit the postulation of the object defined. In general, however, whenever one defines an object, one must in fact prove its existence. "For example, arithmetic assumes the meaning of odd and even, square and cube, geometry that of incommensurable, . . . , whereas the existence of these attributes is demonstrated by means of the axioms and from previous conclusions as premises."[12] Aristotle also listed certain basic principles of argument, principles that earlier thinkers had used intuitively. One such principle is that a given assertion cannot be both true and false. A second principle is that an assertion must be either true or false; there is no other possibility.

For Aristotle, logical argument according to his methods is the only certain way of attaining scientific knowledge. There may be other ways of gaining knowledge, but demonstration via a series of syllogisms is the one way by which one can be sure of the results. Because one cannot prove everything, however, one must always be careful that the premises, or axioms, are true and well known As Aristotle wrote, "syllogism there may indeed be without these conditions, but such syllogism, not being productive of scientific knowledge, will not be demonstration."[13] In other words, one can choose any axioms one wants and draw conclusions from them, but if one wants to attain knowledge, one must start with "true" axioms. The question then becomes, how can one be sure that one's axioms are true? Aristotle answered that these primary premises are learned by induction, by drawing conclusions from our own sense perception of numerous examples. This question of the "truth" of the basic axioms has been discussed by mathematicians and philosophers ever since Aristotle's time. On the other hand, Aristotle's rules of attaining knowledge by beginning with axioms and using demonstrations to gain new results has become the model for mathematicians to the present day.

Although Aristotle emphasized the use of syllogisms as the building blocks of logical arguments, Greek mathematicians apparently never used them. They used other forms, as have most mathematicians down to the present. Why Aristotle therefore insisted on syllogisms is not clear. The basic forms of argument actually used in mathematical proof were analyzed in some detail in the third century BCE by the Stoics, of whom the most prominent was Chrysippus (280–206 BCE). This form of logic is based on **propositions**, statements that can be either true or false, rather than on the Aristotelian syllogisms. The basic rules of inference dealt with by Chrysippus, with their traditional names, are the following, where p, q, and r stand for propositions:

(1) *Modus ponens*
If p, then q.
p.
Therefore, q.

(2) *Modus tollens*
If p, then q.
Not q.
Therefore, not p.

(3) *Hypothetical syllogism*
If p, then q.
If q, then r.
Therefore, if p, then r.

(4) *Alternative syllogism*
p or q.
Not p.
Therefore, q.

For example, from the statements "if it is daytime, then it is light" and "it is daytime," one can conclude by *modus ponens* that "it is light." From "if it is daytime, then it is light" and "it is not light," one concludes by *modus tollens* that "it is not daytime." Adding to the first hypothesis the statement "if it is light, then I can see well," one concludes by the hypothetical syllogism that "if it is daytime, then I can see well." Finally, from "either it is daytime or it is nighttime" and "it is not daytime," the rule of the alternative syllogism allows us to conclude that "it is nighttime."

2.3.2 Number versus Magnitude

Another of Aristotle's contributions was the introduction into mathematics of the distinction between number and magnitude. The Pythagoreans had insisted that all was number, but Aristotle rejected that idea. Although he placed number and magnitude in a single category, "quantity," he divided this category into two classes, the discrete (number) and the continuous (magnitude). As examples of the latter, he cited lines, surfaces, bodies, and time. The primary distinction between these two classes is that a magnitude is "that which is divisible into divisibles that are infinitely divisible,"[14] while the basis of number is the indivisible unit. Thus, magnitudes cannot be composed of indivisible elements, whereas numbers inevitably are.

Aristotle further clarified this idea in his definition of "in succession" and "continuous." Things are **in succession** if there is nothing of their own kind intermediate between them. For example, the numbers 3 and 4 are in succession. Things are **continuous** when they touch and when "the touching limits of each become one and the same."[15] Line segments are therefore continuous if they share an endpoint. Points cannot make up a line, because they would have to be in contact and share a limit. Since points have no parts, this is impossible. It is also impossible for points on a line to be in succession, that is, for there to be a "next point." For between two points on a line is a line segment, and one can always find a point on that segment.

Today, a line segment is considered to be composed of an infinite collection of points, but to Aristotle this would make no sense. He did not conceive of a completed or actual infinity. Although he used the term "infinity," he only considered it as potential. For example, one can bisect a continuous magnitude as often as one wishes, and one can count these bisections. But in neither case does one ever come to an end. Furthermore, mathematicians really do not need infinite quantities such as infinite straight lines. They only need to postulate the existence of, for example, arbitrarily long straight lines.

2.3.3 Zeno's Paradoxes

One of the reasons Aristotle had such an extended discussion of the notions of infinity, indivisibles, continuity, and discreteness was that he wanted to refute the famous paradoxes of Zeno. Zeno stated these paradoxes, perhaps in an attempt to show that the then current notions of motion were not sufficiently clear, but also to show that any way of dividing space or time must lead to problems. The first paradox, the *Dichotomy*, "asserts the non-existence of motion on the ground that that which is in locomotion must arrive at the half-way stage before it arrives at the goal."[16] (Of course, it must then cover the half of the half before it reaches the middle, etc.) The basic contention here is that an object cannot cover a finite distance by moving during an infinite sequence of time intervals. The second paradox, the *Achilles*, asserts a similar point: "In a race, the quickest runner can never overtake the slowest, since the pursuer must first reach the point whence the pursued started, so that the slower must always hold a lead."[17] Aristotle, in refuting the paradoxes, concedes that time, like distance, is infinitely divisible. But he is not bothered by an object covering an infinity of intervals in a finite amount of time. For "while a thing in a finite time cannot come in contact with things quantitatively infinite, it can come in contact with things infinite in respect to divisibility, for in this sense time itself is also infinite."[18] In fact, given the motion in either of these paradoxes, one can calculate when one will reach the goal or when the fastest runner will overtake the slowest.

Zeno's third and fourth paradoxes show what happens when one asserts that a continuous magnitude is composed of indivisible elements. The *Arrow* states that "if everything when it occupies an equal space is at rest, and if that which is in locomotion is always occupying such a space at any moment, the flying arrow is therefore motionless."[19] In other words, if there are such things as indivisible instants, the arrow cannot move during that instant. Since if, in addition, time is composed of nothing but instants, then the moving arrow is always at rest. Aristotle refutes this paradox by noting that not only are there no such things as indivisible instants, but motion itself can only be defined in a period of time. A modern refutation, on the other hand, would deny the first premise because motion is now defined by a limit argument.

The paradox of the *Stadium* supposes that there are three sets of identical objects: the A's at rest, the B's moving to the right past the A's, and the C's moving to the left with equal velocity. Suppose the B's have moved one place to the right and the C's one place to the left, so that B_1, which was originally under A_4, is now under A_5, while C_1, originally under A_5, is now under A_4 (Fig. 2.15). Zeno supposes that the objects are indivisible elements of space and that they move to their new positions in an indivisible unit of time. But since there must have been a moment at which B_1 was directly over C_1, there are two possibilities. Either the two objects did not cross, and so there was no motion at all, or in the indivisible instant, each object had occupied two separate positions, so that the instant was in fact not indivisible. Aristotle believed that he had refuted this paradox because he had already denied the original assumption—that time is composed of indivisible instants.

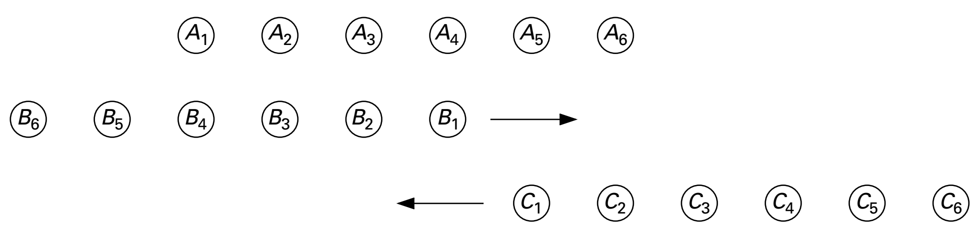


FIGURE 2.15
Zeno's parardox of the *Stadium*

Interestingly, the four paradoxes exhaust the four possibilities of divisibility/indivisibility of space and time. That is, in the *Arrow* both space and time are assumed infinitely divisible, in the *Stadium* both are assumed ultimately indivisible, in the *Dichotomy* space is assumed divisible and time indivisible, and in the *Achilles* the reverse is assumed. So Zeno has shown each of the four possibilities leads to a contradiction.

Controversy regarding these paradoxes has lasted throughout history. The ideas contained in Zeno's statements and Aristotle's attempts at refutation have been extremely fruitful in forcing mathematicians to the present day to think carefully about their assumptions in dealing with the concepts of the infinite or the infinitely small. And in Greek times they were probably a significant factor in the development of the distinction between continuous magnitude and discrete number so important to Aristotle and ultimately to Euclid.

EXERCISES

1. Represent 125, 62, 4821, and 23,855 in the Greek alphabetic notation.

2. Represent 8/9 as a sum of distinct unit fractions. Express the result in the Greek notation. Note that the answer to this problem is not unique.

3. Represent 200/9 as the sum of an integer and distinct unit fractions. Express the result in Greek notation.

4. There are extant Greek land surveys that give measurements of fields and then find the area so the land can be assessed for tax purposes. In general, areas of quadrilateral fields were approximated by multiplying together the averages of the two pairs of opposite sides. In one document, one pair of sides is given as $a = 1/4 + 1/8 + 1/16 + 1/32$ and $c = 1/8 + 1/16$, where the lengths are in fractions of a *schonion*, a measure of approximately 150 feet. The second pair of sides is given as $b = 1/2 + 1/4 + 1/8$ and $d = 1$. Find the average of a and c, the average of b and d, and multiply them together to show that the area of the field is approximately $1/4 + 1/16$ square *schonion*. Note that the taxman has rounded up the exact answer (presumably to collect more taxes).

5. Thales is said to have invented a method of finding distances of ships from shore by use of the angle-side-angle theorem. Here is a possible method: Suppose A is a point on shore and S is a ship (Fig. 2.16). Measure the distance AC along a perpendicular to AC and bisect it at B. Draw CE at right angles to AC and pick point E on it in a straight line with B and S. Show that $\triangle EBC \cong \triangle SBA$ and therefore that $SA = EC$.

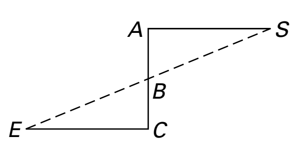


FIGURE 2.16
One method Thales could have used to determine the distance to a ship at sea

6. A second possibility for Thales' method is the following: Suppose Thales was atop a tower on the shore with an instrument made of a straight stick and a crosspiece AC that could be rotated to any desired angle and then would remain where it was put (Fig. 2.17). One rotates AC until one sights the ship S, then turns and sights an object T on shore without moving the crosspiece. Show that $\triangle AET \cong \triangle AES$ and therefore that $SE = ET$.

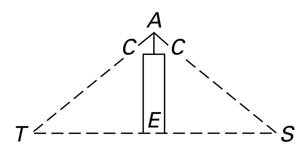


FIGURE 2.17
Second method Thales could have used to determine the distance to a ship at sea

7. Suppose Thales found that at the time a stick of length 6 feet cast a shadow of 9 feet, there was a length of 342 feet from the edge of the pyramid's side to the tip of its shadow. Suppose further that the length of a side of the pyramid was 756 feet. Find the height of the pyramid. (Assuming that the pyramid is laid out so the sides are due north-south and due east-west, this method requires that the sun be exactly in the south when the measurement is taken. When does this occur?[20])

8. Show that the nth triangular number is represented algebraically as $T_n = \frac{n(n+1)}{2}$ and therefore that an oblong number is double a triangular number.

9. Show algebraically that any square number is the sum of two consecutive triangular numbers.

10. Show using dots that eight times any triangular number plus 1 makes a square. Conversely, show that any odd square diminished by 1 becomes eight times a triangular number. Show these results algebraically as well.

11. Show that in a Pythagorean triple, if one of the terms is odd, then two of them must be odd and one even.

12. Construct five Pythagorean triples using the formula $(n, \frac{n^2-1}{2}, \frac{n^2+1}{2})$, where n is odd. Construct five different ones using the formula $(m, (\frac{m}{2})^2 - 1, (\frac{m}{2})^2 + 1)$, where m is even.
13. Show that if a right triangle has one leg of length 1 and a hypotenuse of length 2, then the second leg is incommensurable with the first leg. (In modern terms, this is equivalent to showing that $\sqrt{3}$ is irrational.) Use an argument similar to the proposed Pythagorean argument that the diagonal of a unit square is incommensurable with the side.
14. Show that the areas of similar segments of circles are proportional to the squares on their chords. Assume the result that the areas of circles are proportional to the squares on their diameters.
15. Here is another lune that was "squared" by Hippocrates: Construct a trapezoid $BACD$ such that $BA = AC = CD$ and the square on BD is triple the square on each of the other sides (Fig. 2.18). Then circumscribe a circle around the trapezoid and describe on side BD a circular arc similar to those on the other three sides, that is, an arc whose ratio to side BD is equal to that of the arc on BA to the side BA. Show that the segment on BD is equal to the sum of the segments on BA, AC, and CD. Conclude that the lune bounded by the arcs $BACD$ and BED is equal to the original trapezoid. (Note that you should first prove that the given trapezoid can be constructed and that it can be circumscribed by a circle.)

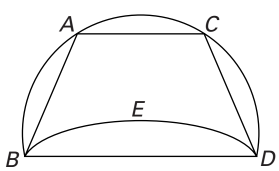


FIGURE 2.18

Hippocrates' lune with outer arc greater than a semicircle

16. Read the entire passage from Plato's *Meno* referred to in the text and write a short essay discussing Socrates' method of convincing the slave boy that he knows how to construct a square double a given square. Consider both the "Socratic method" that Socrates uses as well as the mathematics. (It may be a good idea to do this as a "play" with different students playing the various roles.[21])
17. Consider the quotation from Plato's *Republic:* "It will further our intentions if it [calculation] is pursued for the sake of knowledge and not for commercial ends." Discuss the relevance of this statement to current discussions on the purposes for studying mathematics in school.
18. Give two further examples of each of the two types of syllogisms mentioned in the text.
19. Make up a purposely incorrect syllogism that is related to the correct models in the text. Discuss why its conclusion may be false.
20. Give an example of each of the four rules of inference discussed in the text.
21. In Zeno's *Achilles* paradox, assume the quick runner Achilles is racing against a tortoise. Assume further that the tortoise has a 500-yard head start but that Achilles' speed is fifty times that of the tortoise. Finally, assume that the tortoise moves 1 yard in 5 seconds. Determine the time t it will take until Achilles overtakes the tortoise and the distance d he will have traveled. Note that Achilles must first travel 500 yards to reach the point where the tortoise started. This will take 50 seconds. But in that time the tortoise will move 10 yards farther. Continue this analysis by writing down the sequence of distances that Achilles must travel to reach the point where the tortoise had already been. Show that the sum of this infinite sequence of distances is equal to the distance d calculated first.

REFERENCES AND NOTES

A good source of basic information on Greek civilization is H. D. F. Kitto, *The Greeks* (London: Penguin, 1951). Two excellent general works on early Greek science are by G. E. R. Lloyd, *Early Greek Science: Thales to Aristotle* (New York: Norton, 1970) and *Magic, Reason and Experience* (Cambridge: Cambridge University Press, 1979). The latter work, in particular, deals with the beginnings of logical reasoning in Greece and the emergence of the idea of mathematical proof. The standard reference on Greek mathematics is Thomas Heath, *A History of Greek Mathematics* (New York: Dover, 1981, reprinted from the 1921 original). However, many of Heath's conclusions have been challenged in more recent works. The two best reevaluations of some central parts of the story of Greek mathematics are Wilbur Knorr, *The Ancient Tradition of Geometric Problems* (Boston:

Birkhäuser, 1986), which argues that geometric problem solving was the motivating factor for much of Greek mathematics, and David Fowler's *The Mathematics of Plato's Academy: A New Reconstruction* (Oxford: Clarendon Press, 1987; 2nd edition, 1999), which claims that the idea of *anthyphairesis* (reciprocal subtraction) provides much of the impetus for the Greek development of the ideas of ratio and proportion. A newer work, Serafina Cuomo's *Ancient Mathematics* (London: Routledge, 2001), provides an excellent survey of Greek mathematics, while claiming that many of Heath's (and others') conclusions are based on very flimsy evidence. The emergence of the deductive method in Greek mathematics is discussed in Reviel Netz, *The Shaping of Deduction in Greek Mathematics: A Study in Cognitive History* (Cambridge: Cambridge University Press, 1999). An earlier, but still useful, work on the same topic is I. Mueller, *Philosophy and Deductive Structure in Euclid's Elements* (Cambridge: MIT Press, 1981). Many of the available fragments from the earliest Greek mathematics are collected in Ivor Thomas, *Selections Illustrating the History of Greek Mathematics* (Cambridge: Harvard University Press, 1941).

1. From Proclus's *Summary*, translated in Thomas, *Selections*, I, p. 147.
2. *Plutarch's Moralia*, translated by Phillip H. De Lang and Benedict Einarson (Cambridge: Harvard University Press, 1959), VII, pp. 397–399.
3. See Fowler, *Mathematics of Plato's Academy*, chapter 7, for more on Greek numbers and fractions.
4. Aristotle, *Metaphysics*, 981^b, 14–24. The translations here and below are in the *Great Books* edition (Chicago: Encyclopedia Britannica, 1952), but the references here and to the works of Plato are to lines in the standard Greek text and can be checked in any modern translation.
5. Ibid., 983^a, 14–20.
6. The passage about Socrates and the slave boy is found in Plato, *Meno*, 82^b–85^b.
7. Plato, *Republic* VII, 525. The translation used is that of Frances Cornford.
8. Ibid., VII, 527.
9. Ibid., VII, 528.
10. Ibid., VII, 531.
11. Aristotle, *Prior Analytics* I, 1, 24^b, 19.
12. Aristotle, *Posterior Analytics* I, 10, 76^a, 40–76^b, 10.
13. Ibid., I, 2, 71^b, 23.
14. Aristotle, *Physics* VI, 1, 231^b, 15.
15. Ibid., V, 3, 227^a, 12.
16. Ibid., VI, 9, 239^b, 11. For more on Zeno's paradoxes, see F. Cajori, "History of Zeno's Arguments on Motion," *American Mathematical Monthly* 22 (1915), 1–6, 39–47, 77–82, 109–115, 145–149, 179–186, 215–220, 253–258, 292–297, and H. D. P. Lee, *Zeno of Elea* (Cambridge: Cambridge University Press, 1936).
17. Ibid., VI, 9, 239^b, 15.
18. Ibid., VI, 2, 233^a, 26–29.
19. Ibid., VI, 9, 239^b, 6.
20. For some speculation on how Thales might have accomplished his task, see Lothar Redlin, Ngo Viet, and Saleem Watson, "Thales' Shadow," *Mathematics Magazine* 73 (2000), 347–353.
21. For the text of a large portion of the *Meno* with discussions on how to use this in class, see Victor J. Katz and Karen Michalowicz, eds., *Historical Modules for the Teaching and Learning of Mathematics*, CD (Washington, DC: Mathematical Association of America, 2005), Geometry Module.

CHAPTER 3

Euclid

Not much younger than these [Hermotimus of Colophon and Philippus of Mende, students of Plato] is Euclid, who put together the Elements*, collecting many of Eudoxus's theorems, perfecting many of Theaetetus's, and also bringing to irrefragable demonstration the things which were only somewhat loosely proved by his predecessors. This man lived in the time of the first Ptolemy.*

—Proclus's *Summary* (c. 450 CE) of Eudemus's *History* (c. 320 BCE)[1]

Two legends about Euclid: Ptolemy is said to have asked him if there was any shorter way to geometry than through the *Elements*, and he replied that there was "no royal road to geometry." And, according to Stobaeus (fifth century CE), a student, after learning the first theorem, asked Euclid, "What shall I get by learning these things?" Euclid then asked his slave to give the student a coin, "since he must make gain out of what he learns."[2]

Since the first Ptolemy, Ptolemy I Soter, the Macedonian general of Alexander the Great who became ruler of Egypt after the death of Alexander in 323 BCE and lived until 283 BCE, it is generally assumed from the quotation from Proclus that Euclid flourished around 300 BCE (Fig. 3.1). But besides this date, written down some 750 years later, there is nothing at all known about the life of the author of the *Elements*. Nevertheless, most historians believe that Euclid was one of the first scholars active at the Museum and Library at Alexandria, founded by Ptolemy I and his successor, Ptolemy II Philadelphus. "Museum" here means a "Temple of the Muses," that is, a location where scholars meet and discuss philosophical and literary ideas. The Museum was to be, in effect, a government research establishment. The Fellows of the Museum received stipends and free board and were exempt from taxation. In this way the rulers of Egypt hoped that men of eminence would be attracted there from the entire Greek world. In fact, the Museum and Library soon became a focal point of the highest developments in Greek scholarship, both in the humanities and the sciences. The Fellows were initially appointed to carry on research, but since younger students gathered there as well, the Fellows soon turned to teaching. The aim of the Library was to collect the entire body of Greek literature in the best available copies and to organize it systematically. Ship captains who sailed from Alexandria were instructed to bring back scrolls from every port they touched until their return. The story is told that Ptolemy III, who reigned from 247–221 BCE, borrowed the authorized texts of the playwrights Aeschylus, Sophocles, and Euripides from Athens against a large deposit. But rather than return the originals, he returned only copies. He was quite willing to forfeit the deposit. The Library ultimately contained over 500,000 volumes in every field of knowledge. Although parts of the library were destroyed in various wars, some of it remained intact until the fourth century CE.

FIGURE 3.1

Euclid (detail from Raphael's painting *The School of Athens*). Note that there is no evidence of Euclid's actual appearance.

This chapter will be devoted primarily to a study of Euclid's most important work, the *Elements*, but we will also consider Euclid's *Data*.

3.1 INTRODUCTION TO THE *ELEMENTS*

The *Elements* of Euclid is the most important mathematical text of Greek times and probably of all time. It has appeared in more editions than any work other than the *Bible*. It has been translated into countless languages and has been continuously in print in one country or another nearly since the beginning of printing. Yet to the modern reader the work is incredibly dull. There are no examples; there is no motivation; there are no witty remarks; there is no calculation. There are simply definitions, axioms, theorems, and proofs. Nevertheless, the book has been intensively studied. Biographies of many famous mathematicians indicate that Euclid's work provided their initial introduction into mathematics, that it in fact excited them and motivated them to become mathematicians. It provided them with a model of how "pure mathematics" should be written, with well-thought-out axioms, precise definitions, carefully stated theorems, and logically coherent proofs. Although there were earlier versions of *Elements* before that of Euclid, his is the only one to survive, perhaps because it was the first one written after both the foundations of proportion theory and the theory of irrationals had been developed and the careful distinctions always to be made between number and magnitude had been propounded by Aristotle. It was therefore both "complete" and well organized. Since the mathematical community as a whole was of limited size, once Euclid's work was recognized for its general excellence, there was no reason to keep another inferior work in circulation.

Euclid wrote his text about 2300 years ago. There are, however, no copies of the work dating from that time. The earliest extant fragments include some potsherds discovered in Egypt dating from about 225 BCE, on which are written what appear to be notes on two propositions from Book XIII, and pieces of papyrus containing parts of Book II dating from about 100 BCE. Copies of the work were, however, made regularly from the time of Euclid. Various editors made emendations, added comments, or put in new lemmas. In particular, Theon of Alexandria (fourth century CE) was responsible for one important new edition. Most of the extant manuscripts of Euclid's *Elements* are copies of this edition. The earliest such copy now in existence is in the Bodleian Library at Oxford University and dates from 888. There is, however, one manuscript in the Vatican Library, dating from the tenth century, which is not a copy of Theon's edition but of an earlier version. It was from a detailed comparison of this manuscript with several old manuscript copies of Theon's version that the Danish scholar J. L. Heiberg compiled a definitive Greek version in the 1880s, as close to what he believed the Greek original was as possible. The extracts to be discussed here are all adapted from Thomas Heath's 1908 English translation of Heiberg's Greek. (It should be noted that some modern scholars believe that one can get closer to Euclid's original by taking more account of medieval Arab translations than Heiberg was able to do.)

Euclid's *Elements* is a work in thirteen books. The first six books form a relatively complete treatment of two-dimensional geometric magnitudes while Books VII–IX deal with the theory of numbers, in keeping with Aristotle's instructions to separate the study of magnitude and number. In fact, Euclid included two entirely separate treatments of proportion theory—in Book V for magnitudes and in Book VII for numbers. Book X then provides the link between the two concepts, because it is here that Euclid introduced the notions of commensurability and incommensurability and showed that, with regard to proportions, commensurable magnitudes may be treated as if they were numbers. The book continues by presenting a classification of some incommensurable magnitudes. Euclid next dealt in Book XI with three-dimensional geometric objects and in Book XII with the method of exhaustion applied both to two- and three-dimensional objects. Finally, in Book XIII he constructed the five regular polyhedra and classified some of the lines involved according to his scheme of Book X.

It is useful to note that much of the ancient mathematics discussed in Chapter 1 is included in one form or another in Euclid's masterwork, with the exception of actual methods of arithmetic computation. The methodology, however, is entirely different. Namely, mathematics in earlier cultures always involves numbers and measurement. Numerical algorithms for solving various problems are prominent. The mathematics of Euclid, however, is completely nonarithmetical. There are no numbers used in the entire work aside from a few small positive integers. There is also no measurement. Various geometrical objects are compared, but not by use of numerical measures. There are no cubits or acres or degrees. The only measurement standard—for angles—is the right angle. Nevertheless, the question must be asked as to how much influence the mathematical cultures of Egypt and Mesopotamia had on Euclidean mathematics. In this chapter we discuss certain pieces of evidence in this regard, but a complete answer to this question cannot yet be given.

SIDEBAR 3.1 *Euclid's Postulates and Common Notions*

Postulates

1. To draw a straight line from any point to any point.
2. To produce a finite straight line continuously in a straight line.
3. To describe a circle with any center and distance.
4. That all right angles are equal to one another.
5. That, if a straight line intersecting two straight lines make 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 the angles are less than two right angles.

Common Notions (Axioms)

1. Things which are equal to the same thing are also equal to one another.
2. If equals are added to equals, the wholes are equal.
3. If equals are subtracted from equals, the remainders are equal.
4. Things which coincide with one another are equal to one another.
5. The whole is greater than the part.

3.2 BOOK I AND THE PYTHAGOREAN THEOREM

As Aristotle suggested, a scientific work needs to begin with definitions and axioms. Euclid therefore prefaced several of the thirteen books with definitions of the mathematical objects discussed, most of which are relatively standard. He also prefaced Book I with ten axioms; five of them are geometrical postulates and five are more general truths about mathematics called "common notions." Euclid then proceeded to prove one result after another, each one based on the previous results and/or the axioms. If one reads Book I from the beginning, one never has any idea what will come next. It is only when one gets to the end of the book, where Euclid proved the Pythagorean Theorem, that one realizes that Book I's basic purpose is to lead to the proof of that result. Thus, in order to understand the reasons for various theorems, we begin our discussion of Book I with the Pythagorean Theorem and work backwards. This also enables us to see why certain unproved results must be assumed, namely, the axioms. Sidebar 3.1 does, however, list all of Euclid's axioms (called "postulates" and "common notions") and Sidebar 3.2 has selected definitions.

As we discuss the various propositions, the reader should keep in mind a few important issues. First, although Euclid has modeled the overall structure of the *Elements* using some of Aristotle's ideas, he did not use syllogisms in his proofs. His proofs were written out in natural language and generally used the notions of propositional logic. In fact, one can find examples of all four of the basic rules of inference among Euclid's proofs. Next, Euclid always assumed that if he proved a result for a particular configuration representing the hypotheses of the theorem and illustrated in a diagram, he had proved the result generally. For example, as we will see, he proved the Pythagorean Theorem by drawing some lines and marking some points on a particular right triangle, then arguing to his result on that triangle, and then concluding that the result is true for any right triangle. Of course, when mathematicians today use that strategy, they base it on explicit ideas of mathematical logic. Euclid, in contrast, never discussed his philosophy of proof; he just went ahead and proved

SIDEBAR 3.2 *Selected Definitions from Euclid's* Elements, *Book I*

1. A **point** is that which has no part.
2. A **line** is breadthless length.
3. The extremities of a line are points.
4. A **straight line** is a line which lies evenly with the points on itself.
5. A **surface** is that which has length and breadth only.
6. The extremities of a surface are lines.
7. A **plane surface** is a surface which lies evenly with the straight lines on itself.
8. A **plane angle** is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.
9. And when the lines containing the angle are straight, the angle is called **rectilinear**.
10. When a straight line meeting another straight line makes the adjacent angles equal to one another, each of the equal angles is **right**, and the first straight line is called a **perpendicular** to the second line.
15. A **circle** is a plane figure contained by one line such that all the straight lines meeting it from one point among those lying within the figure are equal to one another.
16. And the point is called the **center** of the circle.
17. A **diameter** of the circle is any straight line drawn through the center and terminated in both directions by the circumference of the circle, and such a straight line also bisects the circle.
18. A **semicircle** is the figure contained by the diameter and the circumference cut off by it. And the center of the semicircle is the same as that of the circle.
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.

things. Of course, occasionally, he seems to have depended on the diagram more than modern mathematicians would allow. These so-called gaps in Euclid's logic were discussed extensively in the nineteenth century, so we will refer to them briefly when they occur here.

We are now ready to state Euclid's version of the Pythagorean Theorem:

PROPOSITION I–47 *In right-angled triangles the square on the hypotenuse is equal to the sum of the squares on the legs.*

Euclid proved the result for triangle ABC by first constructing the line AL parallel to BD meeting the base DE of the square on the hypotenuse at L and then showing that rectangle BL is equal to the square on AB and rectangle CL is equal to the square on AC (Fig. 3.2). To accomplish the first equality, Euclid connected AD and CF to produce triangles ADB and CBF. He then showed that these two triangles are equal to each other, that rectangle BL is double triangle ABD, and that the square on AB is double triangle CBF. His first equality then follows. The second one is proved similarly, while the sum of the two equalities proves the theorem, given common notion 2, that equals added to equals are equal.

We need to understand here what Euclid meant when he claimed that two plane figures are equal. Evidently, he meant that the figures have "equal area," but he nowhere defined this notion, nor did he calculate any areas. His alternative was generally to decompose the regions involved and to show that individual pieces are, in fact, identical. This process is justified by common notion 4, that things that coincide are equal. We will look at this in more detail later. But first, let us see what results we need to make Euclid's proof of I–47 work. First, of course, to make any sense of the theorem at all, we need to know how to construct a square on a given

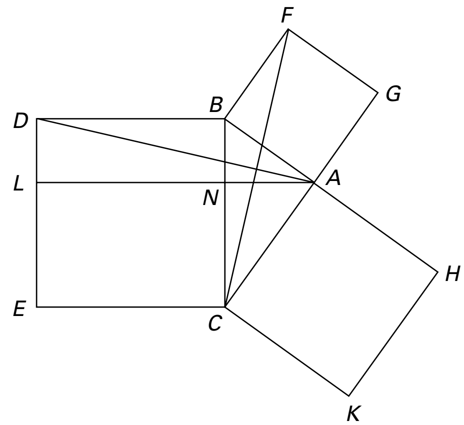


FIGURE 3.2

The Pythagorean Theorem in Euclid's *Elements*

straight line segment. After all, the theorem states a relationship between certain squares. We are therefore led to

PROPOSITION I–46 *On a given straight line to describe a square.*

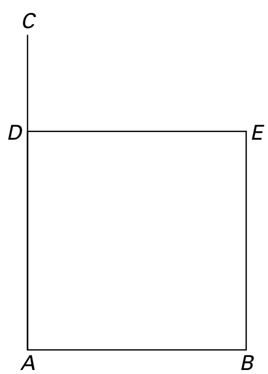


FIGURE 3.3

Elements, Proposition I–46

There are many ways to accomplish this construction, so Euclid had to make a choice. He began by constructing a perpendicular AC to the given line AB and determining a point D so that $AD = AB$. He then constructed a line through D parallel to AB and a line through B parallel to AD, the two lines meeting at point E. His claim now is that quadrilateral $ADEB$ is the desired square (Fig. 3.3). (Note that to get this far we need to be able to construct lines perpendicular and parallel to given lines—these constructions are given in Propositions I–11 and I–31, respectively—as well as cut off on one line segment a line segment equal to another one (Proposition I–3).) To prove that his construction is correct, Euclid began by noting that quadrilateral $ADEB$ has two pairs of parallel sides, so it is a parallelogram. And by Proposition I–34, the opposite sides are equal. It follows that all four sides of $ADEB$ are equal. To show that it is a square, it remains to show that all the angles are right angles. But line AD crosses the two parallel lines AB, DE. So by Proposition I–29, the two interior angles on the same side, namely, angles BAD and ADE, are equal to two right angles. But since we already know that angle BAD is a right angle, so is angle ADE. And since opposite angles in parallelograms are equal according to I–34, all four angles are right, and $ADEB$ is a square.

So although the actual construction of a square is fairly obvious, the proof that the construction is correct appears to require many other propositions. Before looking at some of those propositions, let us return to the main theorem and see what else we need.

The first result is the one that allows Euclid to conclude that triangles ADE and CBF are equal. That follows by the familiar side-angle-side theorem (SAS), proved by Euclid as

PROPOSITION I–4 *If two triangles have two sides equal to two sides respectively, and have the angles contained by the equal sides also equal, then the two triangles are congruent.*

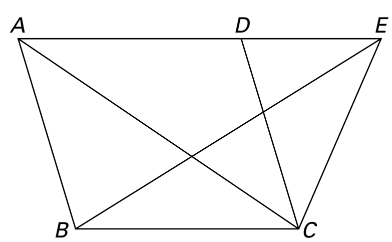
The word "congruent" is used here as a modern shorthand for Euclid's conclusion that each part of one triangle is equal to the corresponding part of the other. Euclid proved this theorem by superposition. Namely, he imagined the first triangle being moved from its original position and placed on the second triangle with one side placed on the corresponding equal side and the angles also matching. Euclid here tacitly assumed that such a motion is always possible without deformation. Rather than supply such a postulate, nineteenth-century mathematicians tended to assume this theorem itself as a postulate.

Euclid also needed the result that a rectangle is double a triangle with the same base and height. This follows from

PROPOSITION I–41 *If a parallelogram has the same base with a triangle and is in the same parallels, the parallelogram is double the triangle.*

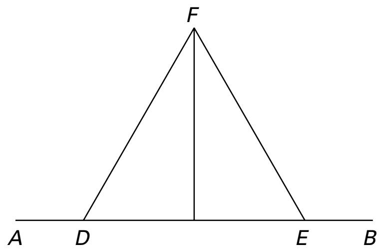
Since "in the same parallels" means from a modern point of view that the two figures have the same height, it would seem that this proposition follows from the formulas for the areas of a triangle and a parallelogram, namely, $A = \frac{1}{2}bh$ and $A = bh$. But, as noted earlier, Euclid did not use formulas to deal with equal area; he used decomposition. So here he showed that the parallelogram can be divided into two triangles, each equal to the given one. In Figure 3.4, the given parallelogram is $ABCD$ and the given triangle is BCE. Euclid drew AC, the diagonal of the parallelogram, then noted that triangle ABC is equal to triangle BCE because they have the same base and are in the same parallels (Proposition I–37). But now parallelogram $ABCD$ is double triangle ABC (by Proposition I–34) and therefore is double triangle BCE.

FIGURE 3.4
Elements, Proposition I–41



Recall that the construction of a square required the construction of both a perpendicular to a given line and a parallel to a given line. The first of these constructions (Proposition I–11) begins with the drawing of the equilateral triangle DFE in which the midpoint C of DE is the point at which the perpendicular is drawn (Fig. 3.5). The construction of an equilateral triangle is accomplished in Proposition I–1, in which Euclid drew circles of radius DE centered on each of the points D and E and then found F as the intersection of the two circles. This construction in turn requires the use of a compass and a straightedge. Namely, Euclid needed to postulate that a circle can be drawn with a given center and radius and that a line can be drawn connecting two points. These postulates are postulate 3 and postulate 1, respectively. But even with these two postulates, modern commentators have noted that there is a logical gap in this proof. How did Euclid know that the two circles drawn from the endpoints of DE actually intersect? It seems obvious in the diagram, but some postulate of continuity is necessary. This was supplied in the nineteenth century and will be discussed later. But once the triangle is constructed, the line from the vertex F to the midpoint C of the base is the desired perpendicular. To prove this, Euclid noted that the two triangles DCF and ECF are congruent by side-side-side (SSS), a result proved as Proposition I–8, by superposition, like

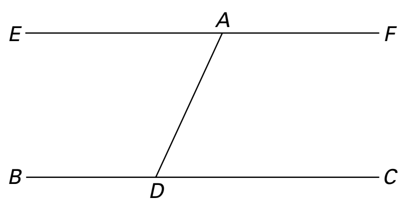
FIGURE 3.5
Elements, Proposition I–11



SAS. Since the sum of the equal angles DCF and ECF is two right angles, each of the angles DCF and ECF is right.

To construct a line through a given point A parallel to a given line BC (Proposition I–31), Euclid took an arbitrary point D on BC and connected AD (Fig. 3.6). By Proposition I–23, he then constructed the angle DAE equal to the angle ADC and extended AE into the straight line AF. That one can extend a straight line in a straight line is the substance of another construction postulate, postulate 2. To prove that EF is now parallel to BC, Euclid noted that the alternate interior angles DAE and ADC are equal. By Proposition I–27, the two lines are parallel.

FIGURE 3.6
Elements, Proposition I–31

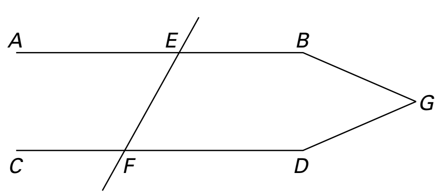


Let us now consider

PROPOSITION I–27 *If a straight line falling on two straight lines makes the alternate angles equal to one another, then the straight lines are parallel to one another.*

Here Euclid argued by *reductio ad absurdum*, a version of *modus tollens*. Namely, he assumed that even though the alternate angles AEF, EFD, formed by line EF falling on lines AB and CD are equal, the lines themselves are not parallel (Fig. 3.7). Therefore, they must meet at point G. It follows that in triangle EFG, the exterior angle AEF equals the interior angle EFD. But this contradicts Proposition I–16, so the original assumption must be false and AB is parallel to CD.

FIGURE 3.7
Elements, Proposition I–27



This then takes us back to

PROPOSITION I–16 *In any triangle, if one of the sides is produced, the exterior angle is greater than either of the interior and opposite angles.*

Suppose side BC of triangle ABC is produced to D (Fig. 3.8). Bisect AC at E and join BE. Euclid then claimed that BE may be extended to F so that $EF = BE$. Unfortunately, there is no postulate allowing him to extend a line to any arbitrary length. Of course, if that assumption is granted, then the proof is straightforward. One connects FC and shows that the triangles ABE and CFE are congruent. Thus, $\angle BAE = \angle ECF$. But $\angle ECF$ is part of the exterior angle ACD; thus, the latter angle is greater than $\angle BAE$. This last statement also requires a postulate, that the whole is always greater than the part (common notion 5).

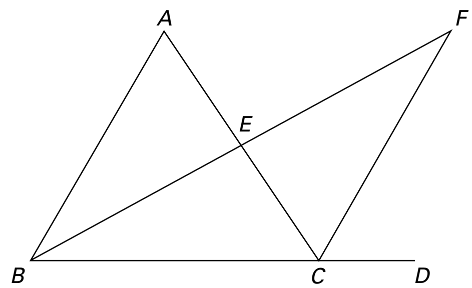


FIGURE 3.8
Elements, Proposition I–16

An immediate corollary is Proposition I–17, that two angles of any triangle are always less than two right angles. As will be discussed later, this proposition, based on the faulty proof of Proposition I–16, was important in the developments leading to the discovery of non-Euclidean geometry.

We could continue by analyzing the proof of I–23, which was used in I–31. This would force us to analyze most of the earlier results in Book I as well. So we will leave some of those results for the exericises and conclude this section by considering just two more important propositions that have already been quoted several times. First, we look at

PROPOSITION I–34 *In parallelograms the opposite sides and angles are equal to one another and the diameter bisects the areas.*

Note that in the proofs of Propositions I–46 and I–41, we have used all three conclusions of this proposition. To prove it, one thinks of the diagonal as first cutting one pair of parallel sides and then cutting the other. In each case, Proposition I–29 implies that the alternate interior angles are equal. It then follows (by angle-side-angle) that the two triangles into which the diagonal cuts the parallelogram are congruent. (The angle-side-angle triangle congruence theorem is Proposition I–26.) The congruence of the two triangles then implies that each pair of opposite sides and each pair of opposite angles are equal. The third part of the proposition follows immediately.

The final proposition we consider is one on which both I–34 and I–46 depend:

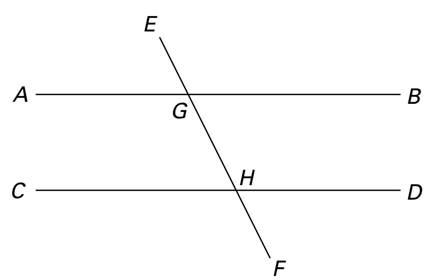
PROPOSITION I–29 *A straight line falling on parallel straight lines makes the alternate angles equal to one another, the exterior angle equal to the interior and opposite angle, and the interior angles on the same side equal to two right angles.*

It is easy enough to see that any two of the statements are simple consequences of the third. So we need to decide which one to prove. From hints in various Greek texts, we know that before Euclid, the situation regarding this theorem was very unclear. How do you prove one of these results? What must you assume? It is in his answer to these questions that Euclid showed his genius. He had already proved the converse of this theorem in Propositions I–27 and I–28. Evidently, however, he saw no way of proving any of the statements in this proposition directly. We can imagine that he struggled with this, but he eventually realized that he would have to take one of these results—or its equivalent—as a postulate. And so he decided, for reasons we cannot guess, to take the contrapositive of the third statement in the proposition as a postulate. Thus, at the beginning of Book I, he placed

POSTULATE 5 *If a straight line intersecting two straight lines make 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 the angles are less than two right angles.*

Given this postulate, the proof of Proposition I–29 is straightforward by a *reductio* argument: Assume that angle AGH is greater than angle GHD (Fig. 3.9). Then the sum of angles AGH and BGH is greater than the sum of angles GHD and BGH. The first sum equals two right angles (by Proposition I–13), so the second one is less than two right angles. Then by the postulate, the lines AB and CD must meet. But this contradicts the hypothesis that those lines are parallel.

FIGURE 3.9
Elements, Proposition I–29



Thus, we see that the Pythagorean Theorem, the culminating theorem of Book I, besides requiring very many of the earlier results in Book I (including all three triangle congruence theorems), rests on the critical parallel postulate. The parallel postulate, alone among Euclid's postulates, has caused immense controversy over the years, because many people felt it was not self-evident. And for Euclid, as for Aristotle, a postulate should be "self-evident." Thus, almost from the time the *Elements* appeared, people have attempted to prove this result as a theorem, using as a basis just the other axioms and postulates. Many people thought they had accomplished this task, but a close examination of every such proof always reveals either an error or, more likely, another assumption—one that perhaps is more self-evident than Euclid's postulate but nevertheless cannot be proved from the other nine axioms. Probably the most familiar "other assumption" is what is now known as

PLAYFAIR'S AXIOM *Through a given point outside a given line, exactly one line may be constructed parallel to the given line.*

We leave it as an exercise that this result is entirely equivalent to Euclid's postulate, at least under the assumption that lines of arbitrary length may be drawn and therefore that Proposition I–16 is true.

3.3 BOOK II AND GEOMETRIC ALGEBRA

Book I of the *Elements*, with its familiar geometric results, was a major component of the Greek mathematician's "toolbox," a set of results that were frequently used in any advanced geometric argument. Book II, on the other hand, is quite different. It deals with the relationships between various rectangles and squares and has no obvious goal. In fact, the propositions in Book II are only infrequently used elsewhere in the *Elements*. Thus, the purpose of Book II has been the subject of much debate among students of Greek mathematics. One interpretation, dating from the late nineteenth century but still common today, is that this book, together with a few propositions in Books I and VI, can best be interpreted as "geometric algebra," the representation of algebraic concepts through geometric figures. In other words, the squares of side length a can be thought of as geometric representations of a^2; rectangles with sides of length a and b can be interpreted as the products ab; and relationships among such objects can be interpreted as equations. Of course, one of the issues in this debate is what one means by the term "algebra." If we think of algebra as meaning the finding of unknown quantities, given certain relationships between those and known quantities, regardless of how these quantities are expressed, then there is certainly algebra in Book II, as well as elsewhere in the *Elements*. It is also easy enough to apply some of Euclid's theorems to the solution of quadratic equations—and this was, in fact, done by medieval Islamic mathematicians. But the majority of scholars today believe that Euclid himself really intended in Book II only to display a relatively coherent body of geometric knowledge that could be used in the proof of further geometric theorems, if not in the *Elements* themselves, then in more advanced Greek mathematics such as the study of conic sections. We shall look at some of the arguments about geometric algebra in what follows.[3]

Euclid began Book II with a definition: *Any rectangle is said to be* **contained** *by the two straight lines forming the right angle.* This definition shows Euclid's geometric usage. The statement does not mean that the area of a rectangle is the product of the length by the width. Euclid never multiplied two lengths together, because he had no way of defining such a process for arbitrary lengths. At various places, he multiplied lengths by numbers (that is, positive integers), but otherwise he only wrote of rectangles contained by two lines. One question then is whether one can interpret Euclid's "rectangle" as meaning a "product."

As an example of Euclid's use of this definition, consider

PROPOSITION II–1 *If there are two straight lines, and one of them is cut into any number of segments whatever, the rectangle contained by the two straight lines is equal to the sum of the rectangles contained by the uncut straight line and each of the segments.*

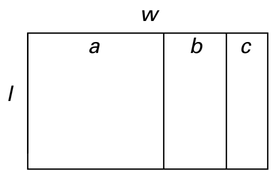


FIGURE 3.10

Elements, Proposition II–1: $l(a + b + c) = la + lb + lc$

We can intepret this algebraically as stating that given a length l and a width w cut into several segments, say, $w = a + b + c$, the area of the rectangle determined by those lines, namely, lw, equals the sum of the areas of the rectangles determined by the length and the segments of the width, namely, $la + lb + lc$. In other words, this theorem states the familiar distributive law: $l(a + b + c) = la + lb + lc$. But let us look more closely at Euclid's proof. Two lines A and BC are given, and the second is divided into three segments by the points D and E (Fig. 3.10). (Euclid had no way of representing "any number" of segments, so he used "three" as what we may call his **generalizable example**.) He then drew BG perpendicular to BC and of length equal to that of A and completed the rectangles $BDKG$, $DELK$, and $ECHL$. Since rectangle $BCHG$ is "the rectangle contained by A and BC," while $BDKG$, $DELK$, and $ECHL$ are the "rectangles contained by A and each of the segments," Euclid could conclude from the diagram that the result was true. At first glance, the proposition seems almost a tautology. But what Euclid seems to be doing here, as well as later in this book, is proving a result about "invisible" figures, that is, the figures stated in the theorem with respect just to the initial two lines and the segments, by using "visible" figures, the actual rectangles drawn. Euclid clearly believed that the "visible" result in the diagram was a correct basis for the proof of the "invisible" result of the proposition.[4] Another example of this process is in

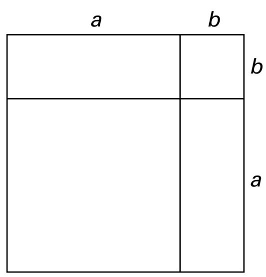


FIGURE 3.11

Elements, Proposition II–4: $(a + b)^2 = a^2 + b^2 + 2ab$

PROPOSITION II–4 *If a straight line is cut at random, the square on the whole is equal to the squares on the segments and twice the rectangle contained by the segments.*

Algebraically, this proposition is simply the rule for squaring a binomial, $(a + b)^2 = a^2 + b^2 + 2ab$, the basis for the square root algorithms discussed in Chapter 1 (Fig. 3.11). Euclid's proof is quite complex, since he needed to prove that the various figures in the diagram are in fact squares and rectangles. But again, he needed to reduce the invisible statement to a visible diagram.

The next two propositions were interpreted in the ninth century CE as geometric justifications of the standard algebraic solutions of quadratic equations.

PROPOSITION II–5 *If a straight line is cut into equal and unequal segments, the rectangle contained by the unequal segments of the whole together with the square on the straight line between the points of section is equal to the square on the half.*

PROPOSITION II–6 *If a straight line is bisected and a straight line is added to it, the rectangle contained by the whole with the added straight line and the added straight line together with the square on the half is equal to the square on the straight line made up of the half and the added straight line.*

Figure 3.12 should help clarify these propositions. If AB is labeled in each diagram as b, AC and BC as $b/2$, and DB as x, Proposition II–5 translates into $(b - x)x + (b/2 - x)^2 = (b/2)^2$, while Proposition II–6 gives $(b + x)x + (b/2)^2 = (b/2 + x)^2$. The quadratic equation $bx - x^2 = c$ [or $(b - x)x = c$] can be solved using the first equality by writing $(b/2 - x)^2 = (b/2)^2 - c$ and then getting

$$x = \frac{b}{2} - \sqrt{\left(\frac{b}{2}\right)^2 - c}.$$

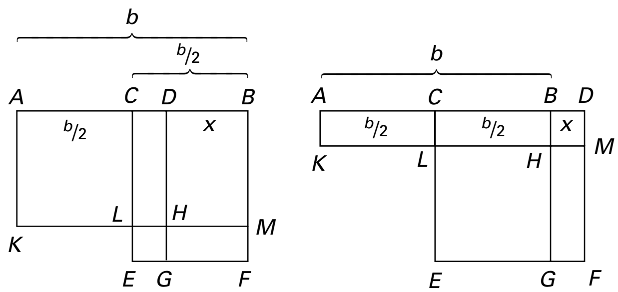


FIGURE 3.12
Elements, Propositions II–5 and II–6

Similarly, the equation $bx + x^2 = c$ (or $(b + x)x = c$) can be solved from the second equality by using an analogous formula. Alternatively, one can label AD as y and DB as x in each diagram and translate the first result into the standard Babylonian system $x + y = b$, $xy = c$, and the second into the system $y - x = b$, $yx = c$. In any case, note that Figure 3.12 is essentially the same as Figure 1.20, the figure representing the Babylonian scribes' probable method for solving the first of these systems.

Euclid, of course, did not do any of the translations indicated. He just used the constructions in Figure 3.12 to prove the equalities of the appropriate squares and rectangles. He did not indicate anywhere that these propositions are of use in solving what we call quadratic equations.

What did these theorems then mean for Euclid? We can see how Proposition II–6 is used in the proof of Proposition II–11, and Proposition II–5 in the proof of Proposition II–14.

PROPOSITION II–11 *To cut a given straight line so that the rectangle contained by the whole and one of the segments is equal to the square on the remaining segment.*

The goal of this proposition is to find a point H on the line so that the rectangle contained by AB and HB equals the square on AH (Fig. 3.13). This is an algebraic problem, in terms of the definition given earlier, since it asks to find an unknown quantity given its relationship to certain known quantities. To translate this problem into modern notation, let the line AB be a and let AH be x. Then $HB = a - x$, and the problem amounts to solving the equation

$$a(a - x) = x^2 \quad \text{or} \quad x^2 + ax = a^2.$$

The Babylonian solution is

$$x = \sqrt{\left(\frac{a}{2}\right)^2 + a^2} - \frac{a}{2}.$$

Euclid's proof seemingly amounts to precisely this formula. To get the square root of the sum of two squares, the obvious method is to use the hypotenuse of a right triangle whose sides are the given roots, in this case, a and $a/2$. So Euclid drew the square on AB and then bisected AC at E. It follows that EB is the desired hypotenuse. To subtract $a/2$ from this length, he drew EF equal to EB and subtracted off AE to get AF; this is the needed value x. Since he wanted the length marked off on AB, he simply chose H so that $AH = AF$. To prove that this choice of H is correct, Euclid then appealed to Proposition II–6: The line AC has been bisected and a straight line AF added to it. Therefore, the rectangle on FC and AF

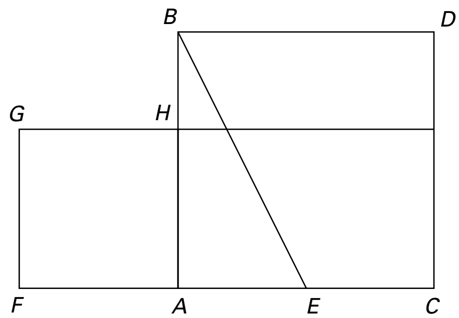


FIGURE 3.13
Elements, Proposition II–11

plus the square on AE equals the square on FE. But the square on FE equals the square on EB, which in turn is the sum of the squares on AE and AB. It follows that the rectangle on FC and AF (equal to the rectangle on FC and FG) equals the square on AB. By subtraction of the common rectangle AK, we get that the square on AH equals the rectangle on HB and AB, as desired.

Euclid has thus solved what we would call a quadratic equation, albeit in geometric dress, in the same manner as the Babylonians. Interestingly enough, he solved the same problem again in the *Elements* as Proposition VI–30. There he wanted to cut a given straight line in "extreme and mean ratio," that is, given a line AB to find a point H such that $AB : AH = AH : HB$. Naturally, this translates algebraically into the same equation as given above. The ratio $a : x$ from that equation, namely, $(\sqrt{5} + 1) : 2$, is generally known as the **golden ratio**. Much has been written about its importance from Greek times to today.[5]

Before considering an example of the use of Proposition II–5, a slight digression back to Book I is necessary.

PROPOSITION I–44 *To a given straight line to apply, in a given rectilinear angle, a parallelogram equal to a given triangle.*

The aim of the construction is to find a parallelogram of given area with one angle given and one side equal to a given line segment. That is, the parallelogram is to be "applied" to the given line segment. This notion of the "application" of areas is, according to some sources, due to the Pythagoreans. That this too can be interpreted algebraically is easily seen if the given angle is a right angle. If the area of the triangle is taken to be c^2 and the given line segment to have length a, the goal of the problem is to find a line segment of length x such that the rectangle with length a and width x has area c^2, that is, to solve the equation $ax = c^2$. Given that Euclid did not deal with "division" of magnitudes, a solution for him amounted to finding the fourth proportional in the proportion $a : c = c : x$. But since he could not use the theory of proportions in Book I, he was forced to use a more complicated method involving areas.

From a geometrical point of view, this construction enables one to compare the sizes of two rectangles. For if rectangle A is applied to one of the sides of rectangle B, then the new rectangle C, equal to A, will share a side with B. Thus, the ratio of the areas of $C = A$ to B will be equal to the ratios of the nonshared sides. Such comparisons, making use of this proposition, are found in the works of Archimedes and Apollonius.

In Proposition I–45, Euclid demonstrated how to construct a rectangle equal to any given rectilinear figure, by simply dividing the figure into triangles and using the result of I–44, among others. This proposition is then used in the first step of the solution of

PROPOSITION II–14 *To construct a square equal to a given rectilinear figure.*

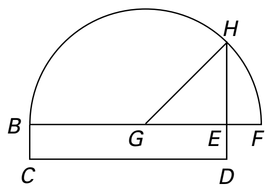


FIGURE 3.14

Elements, Proposition II–14

We can think of this construction as an algebraic problem, since we are asked to find an unknown side of a square meeting certain conditions. In modern notation, we are asked to solve the equation $x^2 = cd$, where c, d are the lengths of the sides of the rectangle constructed, using I–45, equal to the given figure (Fig. 3.14). Placing the sides of the rectangle BE, EF, in a straight line and bisecting BF at G, Euclid constructed the semicircle BHF of radius GF, where H is the intersection of that semicircle with the perpendicular to BF at E. Then, since the straight line BF has been cut into equal segments at G and into unequal segments at E, Proposition II–5 shows that the rectangle contained by BE and EF together with the square on EG is equal to the square on GF. But since $GF = GH$ and the square on GH equals the sum of the squares on GE and EH, it follows that the square on EH satisfies the condition of the problem. Like II–11, Euclid solved this problem a second time using proportions as Proposition VI–13, the construction of a mean proportional between two line segments.

Additionally, in Book VI, Euclid expanded the notion of "application of areas" to applications that are "deficient" or "exceeding." The importance of these notions will be apparent in the discussion of conic sections later. For now, however, we note that in the following two propositions, Euclid solved two types of quadratic equations geometrically.

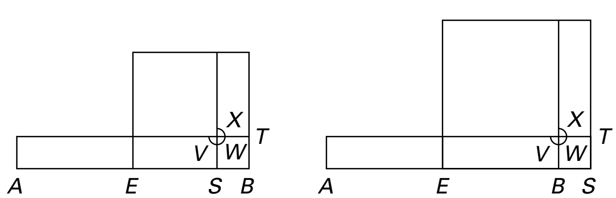
PROPOSITION VI–28 *To a given straight line to apply a parallelogram equal to a given rectilinear figure and deficient by a parallelogram similar to a given one; thus the given rectilinear figure must not be greater than the parallelogram described on the half of the straight line and similar to the defect.*

PROPOSITION VI–29 *To a given straight line to apply a parallelogram equal to a given rectilinear figure and exceeding by a parallelogram figure similar to a given one.*

In the first case, Euclid proposed to construct a parallelogram of given area whose base is less than the given line segment AB. The parallelogram on the deficiency, the line segment SB, is to be similar to a given one. In the second case, the constructed parallelogram of given area has base greater than the given line segment AB, while the parallelogram on the excess, the line segment BS, is again to be similar to a given one (Fig. 3.15). To simplify matters, and to show why we can think of Euclid's constructions as solving quadratic equations, we will assume that the given parallelogram in each case is a square. This implies that the constructed parallelograms are rectangles.

FIGURE 3.15

Elements, Propositions VI–28 and VI–29



Designate AB in both cases by b, and the area of the given rectilinear figure by c. The problems reduce to finding a point S on AB (Proposition VI–28) or on AB extended (Proposition VI–29) so that $x = BS$ satisfies $x(b - x) = c$ in the first case and $x(b + x) = c$ in the second. That is, it is necessary to solve the quadratic equations $bx - x^2 = c$ and $bx + x^2 = c$, respectively. In each case, Euclid found the midpoint E of AB and constructed the square on BE, whose area is $(b/2)^2$. In the first case, S was chosen so that ES is the side of a square whose area is $(b/2)^2 - c$. That is why the condition is stated in the proposition that in effect c cannot be greater than $(b/2)^2$. This choice for ES implies that

$$x = BS = BE - ES = \frac{b}{2} - \sqrt{\left(\frac{b}{2}\right)^2 - c}.$$

In the second case, S was chosen so that ES is the side of a square whose area is $(b/2)^2 + c$. Then

$$x = BS = ES - BE = \sqrt{\left(\frac{b}{2}\right)^2 + c} - \frac{b}{2}.$$

In both cases, Euclid proved that his choice was correct by showing that the desired rectangle equals the gnomon XWV and that the gnomon is in turn equal to the given area c. Algebraically, that amounts in the first case to showing that

$$x(b - x) = \left(\frac{b}{2}\right)^2 - \left[\left(\frac{b}{2}\right)^2 - c\right] = c$$

and in the second that

$$x(b + x) = \left[\left(\frac{b}{2}\right)^2 + c\right] - \left(\frac{b}{2}\right)^2 = c.$$

There has long been a debate over whether the geometric algebra in Euclid stems from a deliberate transformation of the Babylonian quasi-algebraic results into formal geometry. Euclid's solution of several construction problems mirrors the Babylonian solutions of similar problems. One can then argue that the Greek adaptation into their geometric viewpoint, given the necessity of proof, was related to the discovery that not every line segment could be represented by a "number." One can further argue that, once one has translated the material into geometry, one might just as well state and prove certain results for parallelograms as for rectangles, since little extra effort is required. A further argument supporting the transmission and translation is that the original Babylonian methodology itself was couched in a "naive" geometric form, a form well suited to a translation into the more sophisticated Greek geometry.

Was there any opportunity for direct cultural contact between Babylonian mathematical scribes and Greek mathematicians? It used to be argued that this was virtually impossible, because there was no record of Babylonian mathematics at all during the sixth to the fourth centuries BCE, when this contact would have had to take place, and because those in the aristocracy to which the Greek mathematicians belonged would be disdainful of the activities of the scribes, who in Old Babylonian times were not themselves part of the elite. However, recent discoveries have indicated that mathematical activity did continue in the mid-first millennium BCE. Furthermore, by this time, the Mesopotamian languages were often being written in ink on papyrus using a new alphabet. Cuneiform writing on clay tablets was then restricted to important documents that needed to be preserved, and those who could perform this service were now members of the elite, experts in traditional wisdom who were central to the functioning of the state. Besides, from the sixth century BCE on, Mesopotamia was a province of the Persian empire, with whom the Greeks did maintain contact.

On the other hand, despite the possibilities for contact and the logic in the argument of how Babylonian mathematics could have been "translated" into Greek geometry, there is no direct evidence of any transmission of Babylonian mathematics to Greece during or before the fourth century BCE. One could then argue that although the Greeks did employ what we think of as algebraic procedures, their mathematical thought was so geometrical that all such procedures were automatically expressed that way. The Greeks of the period up to 300 BCE had no algebraic notation and therefore no way of manipulating expressions that stood for magnitudes, except by thinking of them in geometric terms. In fact, Greek mathematicians became very proficient in manipulating geometric entities. And finally, we note that there was no way the Greeks could express, other than geometrically, irrational solutions of quadratic equations.

A clear answer to the question of whether Babylonian algebra was transmitted in some form to Greece by the fourth century BCE cannot yet be given. Hopefully, further research in the original sources will enable us to find an answer in the future.

3.4 CIRCLES AND THE PENTAGON CONSTRUCTION

Books I and II dealt with properties of rectilinear figures, that is, figures bounded by straight line segments. In Book III, Euclid turned to the properties of the most fundamental curved figure, the circle. The Greeks were greatly impressed with the symmetry of the circle, the fact that no matter how you turned it, it always appeared the same. They thought of it as the most perfect of plane figures. Similarly, they felt the three-dimensional analogue of the circle, the sphere, was the most perfect of solid figures. These philosophical ideas provided the basis for the Greek ideas on astronomy, which will be discussed in Chapter 5. Many of the theorems in Books III and IV dated from the earliest period of Greek mathematics. As such, they became part of the Greek mathematician's toolbox for solving other problems. As we saw, Hippocrates used results on circles in his quadrature of lunes.

If there is any organizing principle of Book III, it is to provide for the construction, in Book IV, of polygons, both inscribed in and circumscribed about circles. In particular, most of the propositions from the last half of Book III are used in the most difficult construction of Book IV, the construction of the regular pentagon. The constructions of the triangle, square, and hexagon are relatively intuitive and are probably the work of the Pythagoreans. On the other hand, the construction of the pentagon involves more advanced concepts, including the division of a line segment into extreme and mean ratio, and is therefore probably a later development, perhaps due to Theaetetus in the early fourth century BCE. This construction in turn is used in Euclid's construction of some of the regular solids in Book XIII.

SIDEBAR 3.3 *Selected Definitions from Euclid's* Elements, *Book III*

2. A straight line is said to **touch a circle** which, meeting the circle and being produced, does not cut the circle.
6. A **segment of a circle** is the figure contained by a straight line and a circumference of a circle.
8. An **angle in a segment** is the angle which, when a point is taken on the circumference of the segment and straight lines are joined from it to the extremities of the straight line which is the **base of the segment**, is contained by the straight lines so joined.

After presenting a few relevant definitions (Sidebar 3.3), Euclid began Book III with some elementary constructions and propositions, including the very useful result that diameters bisect chords to which they are perpendicular. He then showed how to construct a tangent to a circle:

PROPOSITION III–16 *The straight line drawn at right angles to the diameter of a circle from its extremity will fall outside the circle, and into the space between the straight line and the circumference another straight line cannot be interposed.*

This proposition asserts that the line perpendicular to the diameter at its extremity is what is today called a **tangent**. Euclid only remarked in a corollary that it "touches" the circle, as in definition 2. But the statement that no straight line can be interposed between the curve and the line ultimately became part of the definition of a tangent before the introduction of calculus. Euclid's proof of this result, as to be expected, was by a *reductio* argument.

Propositions III–18 and III–19 give partial converses to Proposition III–16. The former shows that the line from the center of a circle that meets a tangent is perpendicular to the tangent; the latter demonstrates that a perpendicular from the point of contact of a tangent goes through the center of the circle. Propositions III–20 and III–21 also give familiar results, respectively, that the angle at the center is double the angle at the circumference, if both angles cut off the same arc, and that angles in the same segment are equal. The proofs of both are clear from Figure 3.16 as is the proof of Proposition III–22, that the opposite angles of quadrilaterals inscribed in a circle are equal to two right angles.

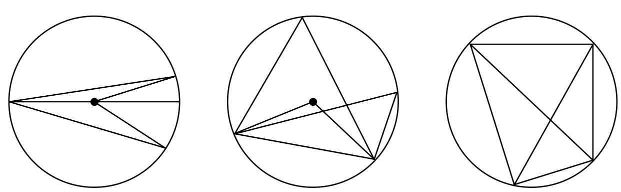


FIGURE 3.16
Elements, Propositions III–20, III–21, and III–22

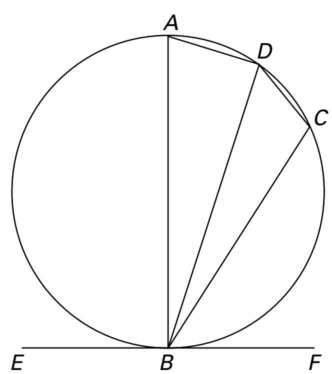
Proposition III–31 asserts that the angle in a semicircle is a right angle. One could conclude this immediately from Proposition III–20, if one is prepared to consider a straight angle as an angle. Then the angle in a semicircle is half of the straight angle of the diameter, which is in turn equal to two right angles. Euclid, however, did not consider a straight angle as an angle, so he gave a different proof.

Proposition III–32 is more complicated, but necessary for the pentagon construction.

PROPOSITION III–32 *If a straight line is tangent to a circle, and from the point of tangency there is drawn a straight line cutting the circle, the angles which that line makes with the tangent will be equal to the angles in the alternate segments of the circle.*

In other words, this proposition asserts that one of the angles formed by the tangent EF and the secant BD, say, angle DBF, is equal to any angle in the "alternate" segment BD of the circle, such as angle DAB (Fig. 3.17). Similarly, the other angle made by the tangent, angle DBE, is equal to any angle in the remaining segment, such as angle DCB. (We can say "any angle" in the segment, since by Proposition III–21, any two angles in the same segment are equal to one another.) To prove this result, we draw a perpendicular AB to the tangent at the point B of tangency. Since a perpendicular to a tangent passes through the center of the circle (Proposition III–19), the angle ADB, being an angle in a semicircle, is a right angle (Proposition III–31). Therefore, angles DAB and ABD sum to a right angle. But angles DBF and ABD also sum to a right angle. It follows that angle DAB equals angle DBF, as claimed. The equality of the other two angles can then be easily established.

FIGURE 3.17
Elements, Proposition III–32



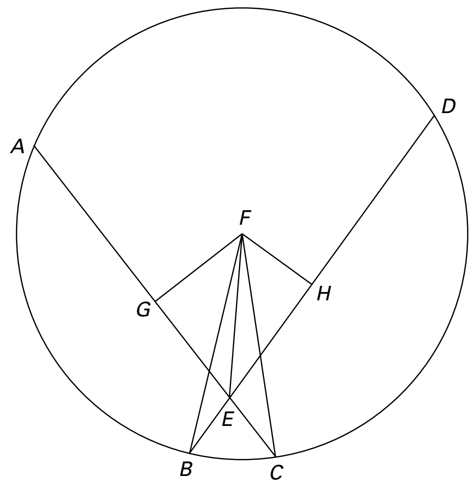
Proposition III–36 is also necessary for the pentagon construction, but because it is closely related to Proposition III–35, and because Propositions II–5 and II–6 make another appearance in these propositions, we first move to

PROPOSITION III–35 *If in a circle two straight lines cut one another, then the rectangle contained by the segments of the one equals the rectangle contained by the segments of the other.*

We note that the rectangles of the proposition are "invisible"; they will only make their appearance through Proposition II–5. For the proof, Euclid first noted that if the two lines meet at the center of the circle, then the result is obvious. Thus, we will assume that the lines AC and BD meet at a point E different from the center F (Fig. 3.18). Draw FG and FH from F perpendicular to AC and DB, and then join FB, FC, and FE. We know that G is then the midpoint of AC. Thus, we can apply II–5 to the line AC and conclude that the rectangle contained by AE and EC together with the square on EG equals the square on GC. By adding the square on GF to both sides and applying the Pythagorean Theorem, we conclude that the rectangle contained by AE and EC plus the square on FE equals the

FIGURE 3.18

Elements, Proposition III–35



square on FC, which in turn equals the square on FB. By the same argument, the rectangle contained by DE and EB plus the square on FE equals the square on FB. It follows that the rectangle contained by DE and EB equals the rectangle contained by AE and EC, as claimed.

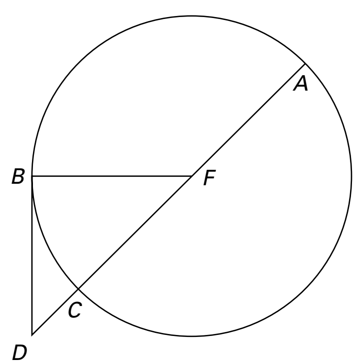
The next proposition deals with two lines cutting the circle that meet outside it:

PROPOSITION III–36 *If from a point outside a circle we draw a tangent and a secant to the circle, then the rectangle contained by the whole secant and that segment which is outside the circle equals the square on the tangent.*

The statement may remind the reader of Proposition II–6. And in fact that proposition is used in the proof. We will just consider the easier case here, where the secant line $DCFA$ goes through the center F (Fig. 3.19). Join FB to form the right triangle FBD. Proposition II–6 now asserts that the rectangle contained by AD and CD, together with the square on FC, equals the square on FD. But $FC = FB$, and the sum of the squares on FB and BD equals the square on FD. Therefore, the rectangle contained by AD and CD equals the square on DB, as claimed. The case where the secant line does not pass through the center is slightly trickier.

FIGURE 3.19

Elements, Proposition III–36



Proposition III–37 is a converse of III–36, asserting that if two straight lines are drawn to a circle from a point outside, one a secant and one touching the circle, and if the relationship between the rectangle and square of that proposition holds, then the second line is a tangent. The proof involves actually drawing a tangent and then showing, using Proposition III–36, that the given line equals the tangent.

The treatment of the pentagon begins in Book IV after Euclid first showed the simpler techniques of inscribing triangles and squares in circles, inscribing circles in triangles and squares, circumscribing triangles and squares about circles, and circumscribing circles about triangles and squares. Euclid then divided his construction of a regular pentagon into two steps, the first being the construction of an isosceles triangle with each of the base angles double the vertex (IV–10), and the second being the actual inscribing of the pentagon in the circle (IV–11). As usual, Euclid did not show how he arrived at the construction, but a close reading of it can well give a clue to his analysis of the problem. We will therefore assume the construction made and try to see where that assumption leads.

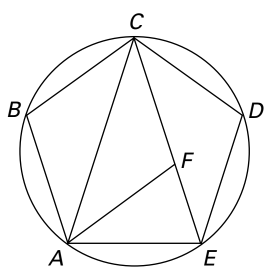
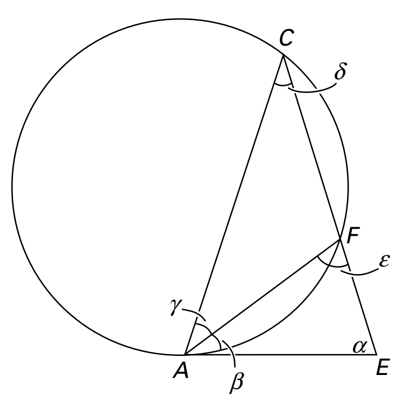


FIGURE 3.20
Construction of a regular pentagon

So suppose $ABCDE$ is a regular pentagon inscribed in a circle (Fig. 3.20). Draw the diagonals AC and CE. Since angles CEA and CAE each subtend an arc double that subtended by angle ACE, it follows that triangle ACE is an isosceles triangle with base angles double those of the vertex. We have therefore reduced the pentagon construction to the construction of that triangle. Assume then that ACE is such an isosceles triangle and let AF bisect angle A. It follows that triangles AFE and CEA are similar, so $EF : AF = EA : CE$. But triangles AFE and AFC are both isosceles, so $EA = AF = FC$. Therefore, $EF : FC = FC : CE$, or, in modern terminology, $FC^2 = EF \cdot CE$. The construction is therefore reduced to finding a point F on a given line segment CE such that the square on CF is equal to the rectangle contained by EF and CE. But this is precisely the construction of Proposition II–11. Once F is found, the isosceles triangle with base angles double the vertex angle can be constructed by drawing a circle centered on C with radius CE and another circle centered on E with radius CF. The intersection A of the two circles is the third vertex of the desired triangle.

Euclid performs this construction in Proposition IV–10 (Fig. 3.21), but could not use similarity arguments in his proof of its validity. He therefore used alternatives. The goal is to show that $\alpha = 2\delta$. If it is shown that $\beta = \delta$, then $\beta + \gamma = \delta + \gamma = \epsilon$. Also, since $\alpha = \beta + \gamma$,

FIGURE 3.21
Elements, Proposition IV–10



then $\epsilon = \alpha$. But then $AE = AF$, and since by construction $AE = FC$, it follows that triangle AFC is isosceles and that $\delta = \gamma$. Finally, $\alpha = \beta + \gamma = \delta + \delta = 2\delta$, as desired. To show that $\beta = \delta$, circumscribe a circle around triangle AFC. Since the rectangle contained by CE, FE, equals the square on FC, it follows that this rectangle also equals the square on AE. Proposition III–37 then asserts that under these conditions on the lines AE and CE, AE is tangent to the circle. Proposition III–32 then allowed Euclid to conclude that $\beta = \delta$ as desired, completing the proof of the construction.

Given the isosceles triangle with base angles double the vertex angle, the inscribing of the regular pentagon in a circle is now straightforward. Euclid first inscribed the isosceles triangle ACE in the circle. Next, he bisected the angles at A and E. The intersection of these bisectors with the circle are points D and B, respectively. Then A, B, C, D, E are the vertices of a regular pentagon.

Euclid completed Book IV with the construction of a regular hexagon and a regular 15-gon in a circle, but did not mention the construction of other regular polygons. Presumably, he was aware that the construction of a polygon of $2^n k$ sides ($k = 3,\ 4,\ 5$) was easy, beginning with the constructions already made, and even that, in analogy with his 15-gon construction, it was straightforward to construct a polygon of kl sides (k, l relatively prime) if one can construct one of k sides as well as one of l sides. Whether he was aware of a construction for the heptagon, however, is not known. In any case, that construction, the first record of which is in the work of Archimedes, would for Euclid be part of advanced mathematics, rather than part of the "elements," because it requires tools other than a straightedge and compass.

3.5 RATIO AND PROPORTION

The regular pentagon is part of the pentagram, evidently one of the symbols used by the Pythagoreans. Thus, it is believed that the Pythagoreans worked out a construction of the pentagon, although more likely their construction used similarity rather than the method described above. It is therefore plausible that the property of the pentagram in reproducing itself when one connects the diagonals of the inner pentagon (Fig. 3.22) could well have been an alternative path to the discovery of incommensurability, rather than the one described earlier. To explain this, we need to move to Book VII, the first of the three books of number theory in the *Elements*.

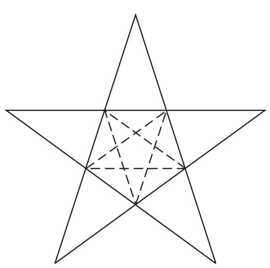


FIGURE 3.22

Diagonals of inner pentagon of a pentagram

Book VII, like all the number theory books, deals with what we call the positive integers in contrast to the geometrical magnitudes of the earlier books. And the first item of business for Euclid here is the familiar process for finding the greatest common divisor of two numbers. This algorithm, usually called the **Euclidean algorithm** although certainly known long before Euclid, is presented in Propositions VII-1 and VII-2. Given two numbers, a, b, with $a > b$, one subtracts b from a as many times as possible; if there is a remainder, c, which of course must be less than b, one then subtracts c from b as many times as possible. Continuing in this manner, one eventually comes either to a number m, which "measures" (divides) the one before (Proposition VII–2), or to the unit (1) (Proposition VII–1). In the first case, Euclid proved that m is the greatest common measure (divisor) of a and b. In the second case, he showed that a and b are prime to one another. For example, given the two numbers 18 and 80, first subtract 18 from 80. One can do this four times, with remainder 8. Next subtract 8

BIOGRAPHY

Theaetetus (417–369 BCE)

Because Plato dedicated a dialogue to him, something is known about Theaetetus's life. He was born near Athens into a wealthy family and was educated there. A meeting with Theodorus of Cyrene before he was 20 excited him about studying mathematics. Theodorus showed him the demonstration that not only was the square root of 2 incommensurable with 1 but so too were the square roots of the other nonsquare integers up to 17. Theaetetus then began research on this issue of incommensurability, both in Heraclea (on the Black Sea) and after 375 BCE in Athens at the Academy. In 369 BCE, he was drafted into the army during a war, was wounded in battle at Corinth, and soon after died of dysentery.

from 18; this can be done twice with remainder 2. Finally, one can subtract 2 exactly four times from 8. It then follows that 2 is the greatest common divisor of 18 and 80. In addition, this calculation shows that one can express the ratio of 80 to 18 in the form (4,2,4), in the sense that the algorithm applied to any other pair a, b, such that $a : b = 80 : 18$, will also give (4,2,4). As another example, take the pair 7 and 32. One can subtract 7 four times from 32 with remainder 4. One can then subtract 4 once from 7 with remainder 3. Finally, one can subtract 3 once from 4 with remainder 1. Thus, 7 and 32 are prime to one another and their ratio can be expressed in the form (4,1,1). (The notation (a,b,c) for ratio is, of course, a modern one.)

It was probably Theaetetus (417–369 BCE) who investigated the possibility of applying the Euclidean algorithm to magnitudes. The results appear as Propositions 2 and 3 of Book X, where we learn how to determine whether two magnitudes A and B have a common measure (are commensurable) or do not (are incommensurable). The procedure, called *anthyphairesis* (reciprocal subtraction), is basically the same as for numbers.[6] Thus, supposing that $A > B$, one first subtracts B from A as many times as possible, say, n_0, getting a remainder b that is less than B. One next subtracts b from B as many times as possible, say, n_1, getting a remainder b_1 less than b. Euclid showed in Proposition X–2 that if this process never ends, then the original two magnitudes are incommensurable. If, on the other hand, one of the magnitudes of this sequence measures the previous one, then that magnitude is the greatest common measure of the original two (Proposition X–3). A natural question here is how one can tell whether or not the process ends. In general, that is difficult. But in certain cases, one observes a repeated pattern in the remainders, which shows that the process cannot end.

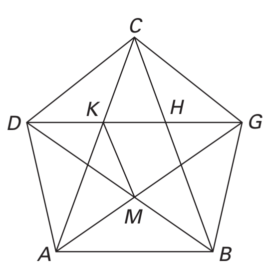


FIGURE 3.23
Incommensurability of diagonal and side of a regular pentagon

For example, let us consider the case of the diagonal and side of the regular pentagon (Fig. 3.23). By the properties of the pentagon, we know that $CG = KG$. Therefore, we can subtract the side $CG = KG$ once from the diagonal GD, leaving remainder KD. We now must subtract KD from the side CG. But $CG = HD$, so KD can be subtracted once from $CG = HD$ with remainder KH. Note that KH is the side of another regular pentagon, whose diagonal is $KM = KD$. Therefore, at the next stage one is again subtracting a side from a diagonal of a pentagon. Since one can continue getting new smaller and smaller pentagons by connecting diagonals of previous ones, it is clear that the process never ends in this case.

BIOGRAPHY

Eudoxus (408–355 BCE)

Eudoxus studied medicine in his youth in Cnidus, an island off the coast of Asia Minor. On a visit to Athens, he was attracted to the lectures at the Academy in philosophy and mathematics and began the study of these subjects. Later, he visited Egypt and was able to make numerous astronomical observations and study the Egyptian calendar. Returning to his home, he opened a school and conducted his own research. Although he returned at least one other time to Athens, this time with his own students, he spent most of the remainder of his life in Cnidus. He is famous not only for his work in geometry but also for his application of spherical geometry to astronomy.

Thus, the diagonal and side of a regular pentagon are incommensurable. In fact, the ratio of the diagonal to the side may be written as (1, 1, 1, . . .).

Given now the existence of incommensurable magnitudes, the Greeks realized that they had to figure out a method of dealing with the ratios of such magnitudes. When they believed that any pair of quantities was commensurable, it was easy enough to see when two such pairs were proportional, or had the same ratio. Euclid in fact defined this concept in Book VII, when he was dealing with numbers: *Four numbers are* **proportional** *when the first is the same multiple, or the same part, or the same parts, of the second that the third is of the fourth*.

As an example, $3 : 4 = 6 : 8$, because 3 is 3 "fourth" parts of 4 while at the same time 6 is 3 "fourth" parts of 8. But for general magnitudes, one cannot use this definition. The side of a pentagon cannot be expressed either as a multiple or as a part or as parts of the diagonal.

Thus, using the anthyphairesis procedure, Theaetetus gave a new definition of "same ratio," which applied to all magnitudes. Suppose there are two pairs of magnitudes A, B, and C, D. Applying this procedure to each pair gives two sequences of equalities:

$$\begin{array}{ll}
A = n_0B + b \ (b < B) & C = m_0D + d \ (d < D) \\
B = n_1b + b_1 \ (b_1 < b) & D = m_1d + d_1 \ (d_1 < d) \\
b = n_2b_1 + b_2 \ (b_2 < b_1) & d = m_2d_1 + d_2 \ (d_2 < d_1) \\
\vdots & \vdots
\end{array}$$

If the two sequences of numbers $(n_0, n_1, n_2, \ldots)$, $(m_0, m_1, m_2, \ldots)$, are equal term by term and both end at, say, $n_k = m_k$, then one can check that the ratios $A : B$ and $C : D$ are both equal to the same ratio of integers. Hence, Theaetetus could give the general definition that $A : B = C : D$ if the (possibly never ending) sequences $(n_0, n_1, n_2, \ldots)$, $(m_0, m_1, m_2, \ldots)$, are equal term by term. Although in general it may be difficult to decide whether two ratios are equal, we have seen that there are interesting cases in which the sequence $n_0, n_1, n_2, \ldots$, is relatively simple to determine. In any case, Aristotle noted that this anthyphairesis definition of equal ratio was the one in use in his time.

Unfortunately, it turned out that Theaetetus's definition was very awkward to use in practice, so the mathematicians continued to search for a better one. It is not known what inspired Eudoxus (408–355 BCE) to give his new definition of same ratio, but a reasonable guess can be made.[7] Theaetetus's definition shows, for example, that if $A : B = C : D$, then

SIDEBAR 3.4 *Selected Definitions from Euclid's* Elements, *Book V*

1. A magnitude is a **part** of a magnitude, the less of the greater, when it measures (divides) the greater.
2. The greater is a **multiple** of the less when it is measured by the less.
3. A **ratio** is a sort of relation in respect to quantity between two magnitudes of the same kind.
4. Magnitudes are said to **have a ratio** to one another which are capable, when multiplied, of exceeding one another.
6. Let magnitudes which have the same ratio be called **proportional**.
7. When, of the equimultiples, the multiple of the first magnitude exceeds the multiple of the second, but the multiple of the third does not exceed the multiple of the fourth, then the first is said to **have a greater ratio** to the second than the third has to the fourth.
9. When three magnitudes are proportional, the first is said to have to the third the **duplicate ratio** of that which it has to the second.
10. When four magnitudes are continuously proportional, the first is said to have to the fourth the **triplicate ratio** of that which it has to the second.

$A > n_0B$ while $C > n_0D$ (since $m_0 = n_0$). Since $n_1A = n_1n_0B + n_1b = (n_1n_0 + 1)B - b_1$, also $n_1A < (n_1n_0 + 1)B$ and similarly $n_1C < (n_1n_0 + 1)D$. A comparison of further multiples of A and B and corresponding multiples of C and D shows that for various pairs r, s, of numbers, $rA > sB$ whenever $rC > sD$ and $rA < sB$ whenever $rC < sD$. Thus, Eudoxus took for his definition of **same ratio** the one now included as definition 5 of Book V (see Sidebar 3.4 for other definitions from Book V):

> 5. Magnitudes are said to be **in the same ratio** (alternatively, **proportional**), the first to the second and the third to the fourth, when, if any equal multiples whatever are taken of the first and third, and any equal multiples whatever of the second and fourth, the former multiples alike exceed, are alike equal to, or alike fall short of, the latter multiples respectively taken in corresponding order.

Translated into algebraic symbolism, this definition says that $a : b = c : d$ if, given any positive integers m, n, whenever $ma > nb$, also $mc > nd$, whenever $ma = nb$, also $mc = nd$, and whenever $ma < nb$, also $mc < nd$. In modern terms, this is equivalent to noting that for every fraction $\frac{n}{m}$, the quotients $\frac{a}{b}$ and $\frac{c}{d}$ are alike greater than, equal to, or less than that fraction.

Of course, before one can define "same ratio," a definition of **ratio** itself is in order. This is given in definitions 3 and 4. Note that Euclid was quite clear that a ratio can only exist between magnitudes of the same kind, that is, lines, surfaces, solids, and so on. In addition, there must be a multiple of each that is greater than the other. So, for example, because no multiple of the angle between the circumference of a circle and a tangent line can exceed a given rectilinear angle, there can be no ratio between these two angles.

Definition 9 is Euclid's version of what is today called the square of a ratio, or, equivalently, the ratio of the squares: If $a : b = b : c$, then $a : c$ is the **duplicate** of the ratio $a : b$. A modern form would be $a : c = (a : b)(b : c) = (a : b)(a : b) = (a : b)^2 = a^2 : b^2$, or, in fractions, $\frac{a}{c} = (\frac{a}{b})^2 = \frac{a^2}{b^2}$. Euclid, however, did not multiply ratios, much less fractions, just as he did not multiply magnitudes. He only multiplied magnitudes by numbers. Similarly, he never divided magnitudes. One cannot interpret Euclid's ratio $a : b$ as a fraction corresponding to a particular point on a number line to which can be applied the standard arithmetical operations. On the other hand, Euclid did use the equivalence between the duplicate ratio of two quantities and the ratio of their squares in the cases where it made sense to speak of the "square" of a quantity (see Proposition VI–20).

The first proposition of Book V asserts, in modern symbols, that if $ma_1, ma_2, \ldots, ma_n$ are equal multiples of $a_1, a_2, \ldots, a_n$, then $ma_1 + ma_2 + \cdots + ma_n = m(a_1 + a_2 + \cdots + a_n)$. Similarly, Proposition V–2 asserts in effect that $ma + na = (m + n)a$, while the next result can be translated as $m(na) = (mn)a$. In other words, these first propositions of Book V give versions of the modern distributive and associative laws.

Proposition V–4 is the first in which the definition of same ratio is invoked. The result states that if $a : b = c : d$, then $ma : nb = mc : nd$, where m, n are arbitrary numbers. To show that equality, Euclid needed to show that if $p(ma)$, $p(mc)$, are equal multiples of ma, mc, and $q(nb)$, $q(nd)$, are equal multiples of nb, nd, then according as $p(ma) >=< q(nb)$, so is $p(mc) >=< q(nd)$. But since $a : b = c : d$, the associative law and the definition of same ratio for the original magnitudes allowed Euclid to conclude the equality of the ratios for the multiples.

The next two propositions repeat the first two with addition being replaced by subtraction. Proposition V–7 shows that if $a = b$, then $a : c = b : c$ and $c : a = c : b$, while Proposition V–8 asserts that if $a > b$, then $a : c > b : c$ and $c : b > c : a$. The proof of the first part of the latter shows Euclid's use of definitions 4 and 7. Since $a > b$, there is an integral multiple, say, m, of $a - b$ that exceeds c (by definition 4). Let q be the first multiple of c that equals or exceeds mb. Then $qc \geq mb > (q - 1)c$. Since $m(a - b) = ma - mb > c$, it follows that $ma > mb + c > qc$. Because also $mb \leq qc$, definition 7 implies that $a : c > b : c$. A similar argument gives the second conclusion.

Among other results of Book V are Proposition V–11, which asserts the transitive law, if $a : b = c : d$ and $c : d = e : f$, then $a : b = e : f$, and Proposition V–16, which states that if $a : b = c : d$, then $a : c = b : d$. The remaining results give other properties of magnitudes in proportion, in particular results dealing with adding or subtracting quantities to the antecedents or consequents in various proportions.

Although Book V gives numerous properties of magnitudes in proportion, the main application of this theory for Euclid was in the treatment of similarity in Book VI. The results of this book then became another major component of the Greek mathematician's toolbox. The book begins with the definition of similarity:

> **Similar rectilinear figures** are such as have their angles respectively equal and the sides about the equal angles proportional.

Recall that the foundation of the idea of similarity, the notion of same ratio (or proportionality), was originally based on the idea that all quantities could be thought of as numbers. So once the basis for the idea of proportionality was destroyed, the foundation for these results no longer existed. That is not to say that mathematicians ceased to use them. Intuitively, they knew that the concept of equal ratio made perfectly good sense, even if they could not provide a formal definition. In Greek times as also in modern times, mathematicians often ignored foundational questions and proceeded to discover new results. The working mathematician knew that eventually the foundation would be strengthened. Once this occurred, the actual similarity results could be organized into a logically acceptable treatise. It is not known who provided this final organization. What is probably true is that there was actually very little to redo except for the proof of the first proposition of the book. That is the only one that depends directly on Eudoxus's definition.

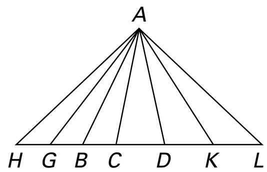


FIGURE 3.24
Elements, Proposition VI–1

PROPOSITION VI–1 *Triangles and parallelograms which have the same height are to one another as their bases.*

Given triangles ABC, ACD, with the same height, Euclid needed to show that as BC is to CD, so is the triangle ABC to the triangle ACD. Proceeding as required by Eudoxus's definition, he extended the base BD to both right and left so that he could take arbitrary multiples of both BC and CD along that line (Fig. 3.24). As earlier, since he could not take an "arbitrary multiple," Euclid used a "generalizable example." So working with two line segments on each side, Euclid noted that because triangles with equal heights and equal bases are equal, whatever multiple the base HC is of the base BC, the triangle AHC is the same multiple of triangle ABC. The same holds for triangle ALC with respect to triangle ACD. Since again triangles AHC and ALC have the same heights, the former is greater than, equal to, or less than the latter precisely when HC is greater than, equal to, or less than CL. Equal multiples having been taken of base BC and triangle ABC, and other equal multiples of base CD and triangle ACD, and the results compared as required by Eudoxus's definition, it follows that $BC : CD = ABC : ACD$ as desired. The result for parallelograms is immediate, because each parallelogram is double the corresponding triangle.

After showing in Proposition VI–2 that a line parallel to one of the sides of a triangle cuts the other two sides proportionally and conversely, and in the following proposition that the bisector of an angle of a triangle cuts the opposite side into segments in the same ratio as that of the remaining sides and conversely, Euclid next gave various conditions under which two triangles are similar. Because the definition of similarity requires both that corresponding angles are equal and that corresponding sides are proportional, Euclid showed that one or the other of these two conditions is sufficient. He also stated the conditions under which the equality of only one pair of angles and the proportionality of two pairs of sides guarantees similarity. Proposition VI–8 then shows that the perpendicular to the hypotenuse from the right angle of a right triangle divides the triangle into two triangles, each similar to the original one.

Among the useful constructions of Book VI are the finding of proportionals. Given line segments a, b, c, Euclid showed how to determine x satisfying $a : b = b : x$ (Proposition VI–11), $a : b = c : x$ (Proposition VI–12), and $a : x = x : b$ (Proposition VI–13). This last result is equivalent to finding a square root, that is, to solving $x^2 = ab$, and is therefore nearly identical to the result of Proposition II–14. In fact, the constructions in the proof are the same; the only difference is that here Euclid used similarity to prove the result, while earlier he used II–5.

Proposition VI–16 is in essence the familiar one that in a proportion the product of the means is equal to the product of the extremes. But since Euclid never multiplied magnitudes, he could not have stated this result in terms of Book V. In the geometry of Book VI, however, he has the equivalent of multiplication, for line segments only:

PROPOSITION VI–16 *If four straight lines are proportional, the rectangle contained by the extremes is equal to the rectangle contained by the means; and if the rectangle contained by the extremes is equal to the rectangle contained by the means, the four straight lines will be proportional.*

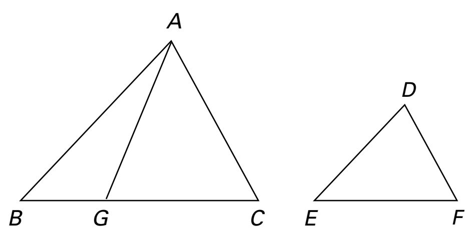
Proposition VI–19 is of fundamental importance later. It also illustrates Euclid's notion of duplicate ratio:

PROPOSITION VI–19 *Similar triangles are to one another in the duplicate ratio of the corresponding sides.*

A modern statement of this result would replace "in the duplicate ratio" by "as the square of the ratio." But Euclid did not multiply either magnitudes or ratios. Ratios are not quantities; they are not to be considered as numbers in any sense of the word. Hence, for this particular proposition, Euclid needed to construct a point G on BC so that $BC : EF = EF : BG$ (Fig. 3.25). The ratio $BC : BG$ is then the duplicate of the ratio $BC : EF$ of the corresponding sides. To prove the result, he showed that the triangles ABG, DEF, are equal. Because triangle ABC is to triangle ABG as BC is to BG, the conclusion follows immediately. Proposition VI–20 extends this result to similar polygons. In particular, the duplicate ratio of two line segments is equal to the ratio of the squares on the segments.

FIGURE 3.25

Elements, Proposition VI–19



Two parallelograms, of course, can be equiangular without being similar. Euclid was also able to deal with the ratio of such figures, but only by using a concept not formally defined:

PROPOSITION VI–23 *Equiangular parallelograms have to one another the ratio compounded of the ratios of the sides.*

The proof shows what Euclid means by the term "compounded," at least in the context of ratios of line segments. If the two ratios are $a : b$ and $c : d$, one first constructs a segment e such that $c : d = b : e$. The ratio **compounded** of $a : b$ and $c : d$ is then the ratio $a : e$. In modern terms, the fraction $\frac{a}{e}$ is simply the product of the fractions $\frac{a}{b}$ and $\frac{c}{d} = \frac{b}{e}$. Interestingly enough, although Euclid never considered compounding again, this notion became quite important in later Greek times as well as in the medieval period.

3.6 NUMBER THEORY

Book VII of the *Elements* is the first of three dealing with the elementary theory of numbers. There is no mention of the first six books in Books VII, VIII, and IX; these three books form an entirely independent unit. Only in later books is there some connection made between the three arithmetic books and the earlier geometric ones. The new start that Euclid made in Book VII is evidence of his desire to stick with Aristotle's clear separation between magnitude and number. The first six books dealt with magnitudes, in particular lengths and areas. The fifth book dealt with the general theory of magnitudes in proportion. But in Books VII–IX

SIDEBAR 3.5 *Selected Definitions from Euclid's* Elements, *Book VII*

1. A **unit** is that by virtue of which each of the things that exist is called one.
2. A **number** is a multitude composed of units.
3. A number is a **part** of a number, the less of the greater, when it measures the greater;
4. but **parts** when it does not measure it.
5. The greater number is a **multiple** of the less when it is measured by the less.
11. A **prime number** is that which is measured by the unit alone.
12. Numbers **prime to one another** are those which are measured by the unit alone as a common measure.
15. A number is said to **multiply** a number when that which is multiplied is added to itself as many times as there are units in the other, and thus some number is produced.
20. Numbers are **proportional** when the first is the same multiple, or the same part, or the same parts, of the second that the third is of the fourth.

Euclid dealt only with numbers. He did not consider these as types of magnitudes, but as entirely separate entities. Therefore, although there are many results in Book VII that appear to be merely special cases of results in Book V, for Euclid they are quite different. One should not be misled by the line segments Euclid used in these books to represent numbers. He did not use the fact of the representation in his proofs. Perhaps this representation was the only one that occurred to him.

It is reasonably certain that many of the propositions in the arithmetic books date back to the Pythagoreans. But from the use of Book VII in Book X, it appears that the details of the compilation of that book are due to the same mathematician who is responsible for Book X, namely, Theaetetus. That is, Theaetetus took the loosely structured number theory of the Pythagoreans and made it rigorous by introducing precise definitions and detailed proofs. It is these that Euclid included in his version of the material.

Book VII, like most of Euclid's books, begins with definitions (Sidebar 3.5). The first definition is, like the beginning definitions of Book I, mathematically useless in modern terms. For Euclid, however, the definition appears as the mathematical abstraction of the concept of "thing." What is more interesting is the second definition, that a number is a **multitude** of units. Since "multitude" means plurality, and the unit is not a plurality, it appears that for Euclid, as for the Pythagoreans earlier, 1 is not a number.

Definitions 3 and 5 are virtual word-for-word repetitions of definitions 1 and 2 in Book V, while definition 4 would make no sense in the context of arbitrary magnitudes. Definitions 11 and 12 are essentially modern definitions of prime and relatively prime, with the note that for Euclid a number does not measure itself. Definition 15 is somewhat curious in that this is the only arithmetic operation defined by Euclid. He assumes that addition and subtraction are known. Note that there is no analogue of this definition in Book V.

Recall that the first two propositions of Book VII deal with the Euclidean algorithm. Several of the next propositions are direct analogues of propositions in Book V. For example, Euclid proved in Propositions VII–5 and VII–6 what amounts to the distributive law $\frac{m}{n}(b+d) = \frac{m}{n}b + \frac{m}{n}d$. He had proved this for magnitudes as Proposition V–1, except that there the result dealt with (integral) multiples rather than the parts—here represented as fractions—of Book VII. Even the proofs of these results are virtually identical. That Euclid did not simply quote results from Book V is evidence that for Euclid number was not a type of magnitude.

Propositions VII–11 through VII–22 include various standard results on numbers in proportion, several of which Euclid proved for magnitudes in Book V. Most are used again in the following two books. In particular, Proposition VII–16 proves the commutativity of multiplication, a nontrivial result given Euclid's definition of multiplication. Proposition VII–19 gives the usual test for proportionality, that $a : b = c : d$ if and only if $ad = bc$. Recall that Euclid had already proved an analogue for line segments (Proposition VI–16). The proof here, however, is quite different. Given that $a : b = c : d$, it follows that $ac : ad = c : d = a : b$. Also $a : b = ac : bc$. Therefore, $ac : ad = ac : bc$. Hence, $ad = bc$. The converse is proved similarly. Proposition VII–20 shows that if a, b, are the smallest numbers in the ratio $a : b$, then a and b each divide c, d, the same number of times, where $c : d = a : b$. It then follows that relatively prime numbers are the least of those in the same ratio and conversely.

Propositions VII–23 through VII–32 deal further with primes and numbers relatively prime to one another. In particular, they present Euclid's theory of divisibility and give, together with Proposition IX–14, a version of the fundamental theorem of arithmetic—that every number can be uniquely expressed as a product of prime numbers.

PROPOSITION VII–31 *Any composite number is measured by some prime number.*

PROPOSITION VII–32 *Any number either is prime or is measured by some prime number.*

The latter proposition is a direct consequence of the former. That one in turn is proved by a technique Euclid used often in the arithmetic books, the least number principle. He began with a composite number a, which is therefore measured (divided) by another number b. If b were prime, the result would follow. If not, then b is in turn measured by c, which will then measure a, and c is in turn either prime or composite. As Euclid then said, "if the investigation is continued in this way, some prime number will be found which will measure the number before it, which will also measure a. For, if it is not found, an infinite series of numbers will measure the number a, each of which is less than the other; which is impossible in numbers." One can again note the distinction between number and magnitude. Any decreasing sequence of numbers has a least element, but the same is not true for magnitudes.

Although Euclid did not do so, it is straightforward to demonstrate from VII–32 that any number can be expressed as the product of prime numbers. To prove that this expression is unique, we need

PROPOSITION VII–30 *If a prime number measures the product of two numbers, it will measure one of them.*

Suppose the prime number p divides ab and p does not divide a. Then $ab = sp$, or $p : a = b : s$. But since p and a are relatively prime, they are the least numbers in that ratio. It follows that b is a multiple of p, or that p divides b. Euclid used this proposition to prove the uniqueness of any prime decomposition in

PROPOSITION IX–14 *If a number is the least of those that are measured by certain prime numbers, then no other prime number will measure it.*

Book VIII primarily deals with numbers in continued proportion, that is, with sequences $a_1, a_2, \dots, a_n$, such that $a_1 : a_2 = a_2 : a_3 = \cdots$. In modern terms, such a sequence is called a geometric progression. It is generally thought today that much of the material in this book is due to Archytas (fifth century BCE), the person from whom Plato received his mathematical training. In particular, Proposition VIII–8 is a generalization of a result due to Archytas and coming out of his interest in music. The original result is that there is no mean proportional between two numbers whose ratio in lowest terms is equal to $(n + 1) : n$. Recall that the ratio of two strings whose sound is an octave apart is 2 : 1. This ratio is the compound of 4 : 3 and 3 : 2, so the octave is composed of a fifth and a fourth. Archytas's result then states that the octave cannot be divided into two equal musical intervals. Of course, in this case, the result is equivalent to the incommensurability of $\sqrt{2}$ with 1. But the result also shows that one cannot divide a whole tone, whose ratio of lengths is 9 : 8, into two equal intervals.

PROPOSITION VIII–8 *If between two numbers there are numbers in continued proportion with them, then, however many numbers are between them in continued proportion, so many will also be in continued proportion between numbers which are in the same ratio as the original numbers.*

Euclid concerned himself in several other propositions of Book VIII with determining the conditions for inserting mean proportional numbers between given numbers of various types. Proposition VIII–11 in particular is the analogue for numbers of a special case of VI–20. Namely, Euclid showed that between two square numbers there is one mean proportional and that the square has to the square the duplicate ratio of that which the side has to the side. Similarly, in Proposition VIII–12, Euclid showed that between two cube numbers there are two mean proportionals and the cube has to the cube the triplicate ratio of that which the side has to the side. This is, of course, the analogue in numbers of Hippocrates' reduction of the problem of doubling the cube to that of finding two mean proportionals.

The final book on number theory is Book IX. Proposition IX–20 shows that there are infinitely many prime numbers:

PROPOSITION IX–20 *Prime numbers are more than any assigned multitude of prime numbers.*

As in earlier proofs, Euclid used the method of generalizable example. He picked just three primes, A, B, C, and showed that one can always find an additional one. To do this, consider the number $N = ABC + 1$. If N is prime, a prime other than those given has been found. If N is composite, then it is divisible by a prime p. Euclid showed that p is distinct from the given primes A, B, C, because none of these divides N. It follows again that a new prime p has been found. Euclid presumably assumed that his readers were convinced that a similar proof will work, no matter how many primes are originally picked.

Propositions IX–21 through IX–34 form a nearly independent unit of very elementary results about even and odd numbers. They probably represent a remnant of the earliest Pythagorean mathematical work. This section includes such results as the sum of even numbers is even, an even sum of odd numbers is even, and an odd sum of odd numbers is odd. These elementary results are followed by two of the most significant results of the entire number theory section of the *Elements*.

PROPOSITION IX–35 *If as many numbers as we please are in continued proportion, and there is subtracted from the second and the last numbers equal to the first, then, as the excess of the second is to the first, so will the excess of the last be to all those before it.*

In effect, this result determines the sum of a geometric progression. Represent the sequence of numbers in "continued proportion" by a, ar, ar^2, ar^3, ..., ar^n, and the sum of "all those before [the last]" by S_n (since there are n terms before ar^n). Euclid's result states that

$$(ar^n - a) : S_n = (ar - a) : a.$$

The modern form for this sum is

$$S_n = \frac{a(r^n - 1)}{r - 1}.$$

The final proposition of Book IX, Proposition IX–36, shows how to find perfect numbers, those that are equal to the sum of all their factors. The result states that if the sum of any number of terms of the sequence 1, 2, 2^2, ..., 2^n is prime, then the product of that sum and 2^n is perfect. For example, $1 + 2 + 2^2 = 7$ is prime; therefore, $7 \times 4 = 28$ is perfect. And, in fact, $28 = 1 + 2 + 4 + 7 + 14$. Other perfect numbers known to the Greeks were 6, corresponding to $1 + 2$; 496, corresponding to $1 + 2 + 4 + 8 + 16$; and 8128, corresponding to $1 + 2 + 4 + 8 + 16 + 32 + 64$. Although several other perfect numbers have been found by using Euclid's criterion, it is still not known whether there are any perfect numbers that do not meet it. Leonhard Euler proved that any even perfect number meets Euclid's criterion, but it is not known whether there are any odd perfect numbers. It is curious, perhaps, that Euclid devoted the culminating theorem of the number theory books to the study of a class of numbers only four of which were known. Nevertheless, the theory of perfect numbers has always proved a fascinating one for mathematicians.

3.7 IRRATIONAL MAGNITUDES

Many historians consider Book X the most important of the *Elements*. It is the longest of the thirteen books and probably the best organized. The purpose of Book X is evidently the classification of certain incommensurable magnitudes. One of the motivations for the book was the desire to characterize the edge lengths of the regular polyhedra, whose construction in Book XIII forms a fitting climax to the *Elements*. Euclid needed a nonnumerical way of comparing the edges of the icosahedron and the dodecahedron to the diameter of the sphere in which they were inscribed. In a manner familiar in modern mathematics, this simple question was to lead to the elaborate classification scheme of Book X, far past its direct answer. Much of this book is attributed to Theaetetus, since he is credited with some of the polyhedral constructions of Book XIII and since it was in Plato's dialogue bearing his name that the question of determining which numbers have square roots incommensurable with the unit was brought up. It is the answer to that question, given early in Book X, that then leads to the general classification.

The introductory definitions give Euclid's understanding of the basic terms "incommensurable" and "irrational" (Sidebar 3.6). The first two definitions are relatively straightforward. The third one, on the other hand, needs some comment. First of all, it includes a theorem, which is proved subsequently in Book X. But secondly, note that Euclid's use of the term "rational" is different from the modern usage. For example, if the assigned straight line has

SIDEBAR 3.6 *Selected Definitions from Euclid's* Elements, *Book X*

1. Those magnitudes are said to be **commensurable** which are measured by the same measure, and those **incommensurable** which cannot have any common measure.
2. Straight lines are **commensurable in square** when the squares on them are measured by the same area, and **incommensurable in square** when the squares on them cannot possibly have any area as a common measure.
3. With these hypotheses, it is proved that there exist straight lines infinite in multitude which are commensurable and incommensurable respectively, some in length only, and others in square also, with an assigned straight line. Let then the assigned straight line be called **rational**, and those straight lines which are commensurable with it, whether in length and in square or in square only, **rational**, but those which are incommensurable with it **irrational**.

length 1, then not only are lines of length $\frac{a}{b}$ called rational, but also lines of length $\sqrt{\frac{a}{b}}$ (where a and b are positive integers).

The first proposition of Book X is fundamental, not only in that book but also in Book XII.

PROPOSITION X–1 *Two unequal magnitudes being given, if from the greater there is subtracted a magnitude greater than its half, and from that which is left a magnitude greater than its half, and if this process is repeated continually, there will be left some magnitude less than the lesser of the given magnitudes.*

The result depends on definition 4 of Book V, the criterion that two given magnitudes have a ratio. That definition requires that some multiple n of the lesser magnitude exceeds the greater. Then n subtractions of magnitudes greater than half of what is left at any stage gives the desired result.

Propositions X–2 and X–3 are the results on anthyphairesis discussed earlier. But since Euclid used the same procedure for magnitudes as he did for numbers in Book VII, he could now connect these two distinct concepts. Namely, Euclid showed in Propositions X–5 and X–6 that magnitudes are commensurable precisely when their ratio is that of a number to a number. So even though number and magnitude are distinct notions, one can now apply the machinery of numerical proportion theory to commensurable magnitudes. The more complicated Eudoxian definition is then only necessary for incommensurable magnitudes.

Proposition X–9 is the result attributed to Theaetetus that provides the generalization of the Pythagorean discovery of the incommensurability of the diagonal of a square with its side, or, in modern terms, of the irrationality of $\sqrt{2}$. Namely, Euclid showed here in effect that the square root of every nonsquare integer is incommensurable with the unit. In Euclid's terminology, the theorem states that two sides of squares are commensurable in length if and only if the squares have the ratio of a square number to a square number. The more interesting part is the "only if" part. Suppose the two sides a, b, are commensurable in length. Then $a : b = c : d$ where c, d, are numbers. Hence, the duplicates of each ratio are equal. But Euclid already showed (VI–20) that the square on a is to the square on b in the duplicate ratio of a to b as well as (VIII–11) that c^2 is to d^2 in the duplicate ratio of c to d. The result then follows.

After some further preliminaries on criteria for incommensurability, Euclid proceeded to the major task of Book X, the classification of certain irrational lengths, lengths that are neither commensurable with a fixed unit length nor commensurable in square with it. The entire classification is too long to discuss here, so only a few of the definitions, those that are of use in Book XIII, will be mentioned to provide some of the flavor of this section. It is significant to note that although each of these irrational lengths can be expressed today as a solution of a polynomial equation, Euclid did not use any algebraic machinery. Everything is done geometrically. Nevertheless, for ease of understanding, numerical examples of each definition are presented.

A **medial** straight line is one that is the side of a square equal to the rectangle contained by two rational straight lines commensurable in square only. For example, because the lengths $1, \sqrt{5}$, are commensurable in square only, and because the rectangle contained by these two lengths has an area equal to $\sqrt{5}$, the length equal to $\sqrt[4]{5}$ is medial. A **binomial** straight line is the sum of two rational straight lines commensurable in square only. So the length $1+\sqrt{5}$ is a binomial. Similarly, the difference of two rational straight lines commensurable in square only is called an **apotome**. The length $\sqrt{5}-1$ provides a simple example. A final, more complicated example is given by Euclid's definition of a **minor** straight line. Such a line is the difference $x-y$ between two straight lines such that x, y, are incommensurable in square, such that x^2+y^2 is rational, and such that xy is a medial area, that is, equal to the square on a medial straight line. For example, if $x=\sqrt{5+2\sqrt{5}}$ and $y=\sqrt{5-2\sqrt{5}}$, then $x-y$ is a minor.

3.8 SOLID GEOMETRY AND THE METHOD OF EXHAUSTION

Book XI of the *Elements* is the first of three books dealing with solid geometry. This book contains the three-dimensional analogues of many of the two-dimensional results of Books I and VI. The introductory definitions include such notions as pyramids, prisms, and cones (Sidebar 3.7). The only definition that is somewhat unusual is that of a sphere, which is defined not by analogy to the definition of a circle but in terms of the rotation of a semicircle about its diameter. Presumably, Euclid used this definition because he did not intend to discuss the properties of a sphere as he had discussed the properties of a circle in Book III. The elementary properties of the sphere were in fact known in Euclid's time and dealt with in other texts, including one due to Euclid himself. In the *Elements*, however, Euclid considered spheres only in Book XII, where he dealt with the volume, and in Book XIII, where he constructed the regular polyhedra and showed how they fit into the sphere. His constructions in Book XIII, in fact, show how these polyhedra are inscribed in a sphere by rotating a semicircle around them, as in his definition.

The propositions of Book XI include some constructions analogous to those of Book I. For example, Proposition XI–11 shows how to draw a straight line perpendicular to a given plane from a point outside it, whereas Proposition XI–12 shows how to draw such a line from a point in the plane. There is also a series of theorems on parallelepipeds. In particular, by analogy with Proposition I–36, Euclid showed that parallelepipeds on equal bases and with the same height are equal (Proposition XI–31), and then, in analogy with VI–1, that parallelepipeds of the same height are to one another as their bases (Proposition XI–32). Also, in analogy

SIDEBAR 3.7 *Selected Definitions from Euclid's* Elements*, Book XI*

12. A **pyramid** is a solid figure, contained by planes, which is constructed from one plane to one point.
13. A **prism** is a solid figure contained by planes two of which, namely those which are opposite, are equal, similar and parallel, while the rest are parallelograms.
14. When, the diameter of a semicircle remaining fixed, the semicircle is carried round and restored again to the same position from which it began to be moved, the figure so comprehended is a **sphere**.
18. When, one leg of a right triangle remaining fixed, the triangle is carried around and restored again to the same position from which it began to be moved, the figure so comprehended is a **cone**. And if the fixed leg is equal to the other leg, the cone will be **right-angled**; if less, **obtuse-angled**; and if greater, **acute-angled**.

with VI–19 and VI–20, he showed in Proposition XI–33 that similar parallelepipeds are to one another in the triplicate ratio of their sides. Hence, the volumes of two similar parallelepipeds are in the ratio of the cubes of any pair of corresponding sides. And in Proposition XI–34, in partial analogy with VI–14 and VI–16, he demonstrated that in equal parallelepipeds, the bases are reciprocally proportional to the heights and conversely. As before, Euclid computed no volumes. Nevertheless, one can easily derive from these theorems the basic results on volumes of parallelepipeds. The "formulas" for volumes of other solids are included in Book XII.

The central feature of Book XII, which distinguishes it from the other books of the *Elements*, is the use of a limiting process, generally known as the method of exhaustion. This process, developed by Eudoxus, is used to deal with the area of a circle as well as the volumes of pyramids, cones, and spheres. "Formulas" giving some of these areas and volumes were known much earlier, but for the Greeks a proof was necessary, and Eudoxus's method provided a proof. What it did not provide was a way of discovering the formulas to begin with.

The main results of Book XII are the following:

PROPOSITION XII–2 *Circles are to one another as the squares on the diameters.*

PROPOSITION XII–7 (COROLLARY) *Any pyramid is a third part of the prism which has the same base with it and equal height.*

PROPOSITION XII–10 *Any cone is a third part of the cylinder which has the same base with it and equal height.*

PROPOSITION XII–18 *Spheres are to one another in the triplicate ratio of their respective diameters.*

The first of these results is Euclid's version of the ancient result on the area of a circle, a version already known to Hippocrates 150 years earlier. In modern terms, it states that the area of a circle is proportional to the square on the diameter. It does not state what the constant of proportionality is, but the proof does provide a method for approximating this. Proposition XII–1, that similar polygons inscribed in circles are to one another as the squares on the diameters, serves as a lemma to this proof. This result in turn is a generalization of the result of VI–20 that similar polygons are to one another in the duplicate ratio of the corresponding sides. It is not difficult to show first of all that one can take any corresponding lines in place of the "corresponding sides," even the diameter of the circle, and secondly that one can replace "duplicate ratio" by "squares."

The main idea of the proof of XII–2 is to "exhaust" the area of a particular circle by inscribing in it polygons of increasingly many sides. In particular, Euclid showed that one can inscribe in the given circle a polygon whose area differs from that of the circle by less than any given area. His proof of the theorem began by assuming that the result is not true. That is, if the two circles C_1, C_2, have areas A_1, A_2, respectively, and diameters d_1, d_2, he assumed that $A_1 : A_2 \neq d_1^2 : d_2^2$. Therefore, there is some area S, either greater or less than A_2, such that $d_1^2 : d_2^2 = A_1 : S$. (Note that Euclid has never proved the existence of a fourth proportional to three arbitrary magnitudes, but only to three lengths. This is therefore another unproved result in Euclid. Its truth needs to come from some kind of continuity argument, but perhaps Euclid ignored it because he did not require the actual construction of such a magnitude.)

Suppose first that $S < A_2$ (Fig. 3.26). Then beginning with an inscribed square and continually bisecting the subtended arcs, inscribe in C_2 a polygon P_2 such that $A_2 > P_2 > S$. In other words, P_2 is to differ from A_2 by less than the difference between A_2 and S. This construction is possible by Proposition X–1, since at each bisection one is increasing the area of the polygon by more than half of the difference between the circle and the polygon. Next inscribe a polygon P_1 in C_1 similar to P_2. By Proposition XII–1, $d_1^2 : d_2^2 = P_1 : P_2$. By assumption, this ratio is also equal to $A_1 : S$. Therefore, $P_1 : A_1 = P_2 : S$. But clearly, $A_1 > P_1$. It follows that $S > P_2$, contradicting the assumption that $S < P_2$. Therefore, S cannot be less than A_2. Euclid proved that S also is not greater than A_2 by reducing it to the case already dealt with. It then follows that the ratio of the circles must be equal to the ratio of the squares on the diameters, as asserted.

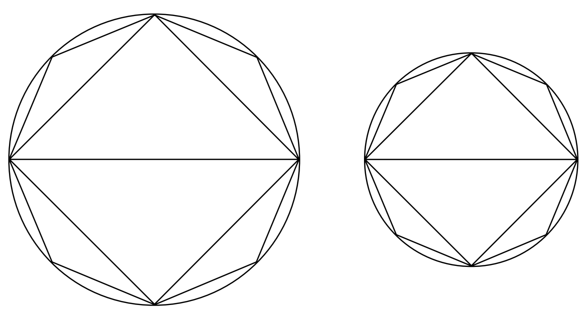


FIGURE 3.26

Elements, Proposition XII–2, the method of exhaustion

FIGURE 3.27

Democritus on a Greek stamp

It is virtually certain that the theorem giving the volume of the pyramid was known to both the Egyptians and the Babylonians (Sidebar 3.8). Archimedes, however, wrote that although Eudoxus was the first to prove that theorem, the result was first discovered by Democritus (fifth century BCE) (Fig. 3.27). Unfortunately, we have no record of how the Egyptians, the Babylonians, or Democritus may have made their discovery. For the latter, we do have a hint in a report given by Chrysippus, in which Democritus discussed the problem of slicing a cone into "indivisible" sections by planes parallel to the base. He wondered whether these indivisible circles would be unequal or equal: "If they are unequal, they will make the cone

SIDEBAR 3.8 *What Did the Greeks Learn from the Egyptians?*

Did the Greeks learn any mathematics from the Egyptians, or was their idea of mathematics so different from that of their predecessors that we may as well assume that they started from scratch? This question has been posed over the years, but because there is no extant documentation of transmission from Egypt to Greece before the third century BCE, we cannot give a definitive answer. Nevertheless, there are certainly hints.

The Greeks in general stated that they had learned from Egypt. The stories the Greeks told about many of their mathematicians, including Pythagoras, Thales, and Eudoxus, note that they studied in Egypt. And many Greek documents say that geometry was first invented by the Egyptians and then passed on to the Greeks. But what is meant here by geometry? It clearly cannot mean an axiomatic treatment such as we find in Euclid's *Elements*. What it could mean, however, is the results themselves. After all, one does not *discover* results by the axiomatic method. One discovers them by experiment, by trial and error, by induction; only after the discovery is made does one worry about actually proving that what one has proposed is correct. So it seems clear that what the Greek writers meant about the Egyptians inventing geometry was the results, not the method of proof. It also seems clear that the idea of proof from a system of axioms is original to the Greeks.

What geometric results could the Greeks have learned? One answer seems to be most of the formulas concerned with the measurement of geometric objects, such as the volume of a pyramid, the area of a circle, and the area of a hemisphere. They could also have learned the basic principles of similarity, since Egyptian sources reveal highly developed proportional thinking connected with the use of scale models. And we are certain that the Greeks learned the use of unit fractions from the Egyptians, although these did not appear in formal Greek mathematics.

Just as in the case of the Babylonians, there is no documentary evidence of direct Egyptian influence on Greek mathematics, but the circumstantial evidence is relatively strong. And as in the case of Babylonian influence, we will have to await further research to answer the question.

There has been much recent historical controversy over the relationship of Greek civilization to Egyptian civilization and, in particular, of the relationship of Greek mathematics to Egyptian mathematics. The opening shot in this battle was the publication of Martin Bernal's *Black Athena: The Afroasiatic Roots of Classical Civilization* (New Brunswick: Rutgers University Press, 1987). This work asserted that classical Greek civilization has deep roots in Afroasiatic cultures, but that these influences have been systematically ignored or denied since the eighteenth century, chiefly for racist reasons. Bernal did not write much about science in this work, but summarized his views on the contributions of Egyptian science to Greek science in "Animadversions on the Origins of Western Science," *Isis* 83 (1992), 596–607. This article was answered by Robert Palter in his "*Black Athena*, Afro-Centrism, and the History of Science," *History of Science* 31 (1993), 227–287. Bernal responded in "Response to Robert Palter," *History of Science* 32 (1994), 445–464; and Palter answered Bernal in the same issue on pages 464–468. The last word on this issue has not yet been uttered.

---

irregular, as having many indentations, like steps, and unevennesses; but if they are equal, the sections will be equal, and the cone will appear to have the property of the cylinder, and to be made up of equal, not unequal, circles, which is very absurd."[8]

Although we do not know what Democritus's final conclusion was, he evidently did think that the cone and, analogously, the pyramid were "made up" of indivisibles. If so, he could have derived Euclid's Proposition XII–5, that pyramids of the same height and with triangular bases are to one another as their bases. For if one imagines the two pyramids cut respectively by planes parallel to and at equal distances from the bases, then the corresponding sections of the two pyramids would be in the ratio of the bases. Since Democritus conceived of each pyramid as being "made up" of these infinitely many indivisible sections, the pyramids themselves would be in this same ratio. He could then have completed the demonstration of the volume formula by noting, as in XII–7, that a prism with a triangular base can be divided into three pyramids, all of equal height and equal bases.

Euclid, of course, proved XII–5 as well as XII–10 and XII–18 by using *reductio* arguments. Assuming the falsity of the given assertion, he proceeded to construct inside the given solid other solids, whose properties are already known, such that the difference between the given solid and the constructed one is less than a given "small" solid, the "error" defined by the false assumption. That is, he exhausted the solid. The known properties of the constructed figure then led him to a contradiction as in the proof of XII–2. But the quotation from Democritus shows us that from the earliest period of Greek mathematics there were attempts to discover certain results by the use of infinitesimals, even though, as we have seen, Aristotle banned such notions from formal Greek mathematics.

The final book of the *Elements*, Book XIII, is devoted to the construction of the five regular polyhedra and their "comprehension" in a sphere (Fig. 3.28). This book is the three-dimensional analogue to Book IV. The study of the five regular polyhedra—the cube, tetrahedron, octahedron, dodecahedron, and icosahedron—and the proof that these are the only regular polyhedra are due to Theaetetus. The first three solids were known in pre-Greek times, and there is archaeological evidence of bronze dodecahedra dating back perhaps to the seventh century BCE. The icosahedron, however, was evidently first studied by Theaetetus. It was also he who recognized that these five were the only regular polyhedra, and that in fact the properties of the regular polyhedra were something to study.

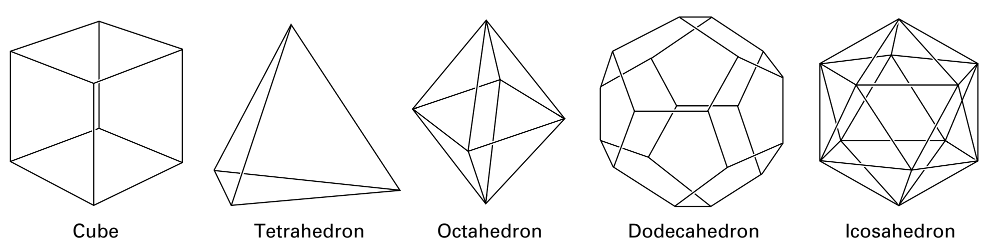


FIGURE 3.28

The five regular polyhedra

Euclid proceeded systematically in Book XIII to construct each of the polyhedra, to demonstrate that each may be comprehended (inscribed) in a sphere, and to compare the edge length of the polyhedron with the diameter of the sphere. For the tetrahedron, Euclid showed that the square on the diameter is $1\frac{1}{2}$ times the square on the edge. In the cube the square on the diameter is triple the square on the edge, whereas in the octahedron the square on the diameter is double that on the edge. The other two cases are somewhat trickier. Euclid proved that the edge of the dodecahedron is an apotome equal in length to the greater segment of the edge of the inscribed cube when that edge is cut in extreme and mean ratio. Thus, if the diameter of the sphere is 1, then the edge of the cube is $c = \frac{\sqrt{3}}{3}$. Therefore, the edge length of the dodecahedron is the positive root of $x^2 + cx - c^2 = 0$ or $\frac{c}{2}(\sqrt{5} - 1) = \frac{1}{6}(\sqrt{15} - \sqrt{3})$. Because both $\sqrt{15}$ and $\sqrt{3}$ are rational by Euclid's definition, and because they are commensurable in square only, the edge length is in fact an apotome.

For the icosahedron, Euclid proved that the side is a minor straight line. In this case, the square on the diameter of the sphere is five times the square on the radius r of the circle circumscribing the five upper triangles of the icosahedron. The bases of these five triangles form a regular pentagon, each edge of which is an edge of the icosahedron. The side of a pentagon inscribed in a circle of radius r is equal to

$$\frac{r}{2}\sqrt{5+2\sqrt{5}} - \frac{r}{2}\sqrt{5-2\sqrt{5}} = \frac{r}{2}\sqrt{10-2\sqrt{5}}.$$

If the diameter of the sphere is 1, then $r = \frac{\sqrt{5}}{5}$, a rational value, and the edge length of the icosahedron is indeed a minor straight line. In particular, this edge length is

$$\frac{\sqrt{5}}{10}\sqrt{10-2\sqrt{5}} = \frac{1}{10}\sqrt{50-10\sqrt{5}}.$$

In a fitting conclusion to Book XIII and the *Elements*, Euclid constructed the edges of the five regular solids in one plane figure, thereby comparing them to each other and the diameter of the given sphere. He then demonstrated that there are no regular polyhedra other than these five.

3.9 EUCLID'S *DATA*

Euclid wrote several mathematics books more advanced than the *Elements*. The most important of the ones that have survived is the *Data*. This was in effect a supplement to Books I–VI of the *Elements*. Each proposition of the *Data* takes certain parts of a geometric configuration as given, or known, and shows that therefore certain other parts are determined. ("Data" means "given" in Latin.) Generally, in his proofs, Euclid showed that these other parts were determined by showing exactly how to determine them. Thus, the *Data* in essence transformed the synthetic purity of the *Elements* into a manual appropriate to one of the goals of Greek mathematics, the solution of new problems.

As one example, consider

PROPOSITION 39 *If each of the sides of a triangle is given in magnitude, the triangle is given in form.*

In other words, this proposition claims that if the lengths of the three sides of a triangle are known, then the triangle itself is determined, that is, not only are the sides known but also the angles. In the demonstration, Euclid carefully constructed a triangle with sides equal to those of the given triangle. He then used parts of the "toolbox," in this case Proposition I–8 and definition 1 of Book VI of the *Elements*, to conclude that the constructed triangle was "equal and similar" to the given triangle. This means, then, that the original triangle was "given in form."

We can certainly consider several of the propositions of the *Data* as examples of geometric algebra, in that Euclid showed how to find unknown lengths, given certain known ones. For example, here are two propositions closely related to *Elements* VI–29.

PROPOSITION 84 *If two straight lines contain a given area in a given angle, and one of them is greater than the other by a given straight line, each of them will be given, too.*

If, as in the discussion of VI–29, it is assumed that the given angle is a right angle—and the diagram in the medieval manuscripts that survive shows such an angle—the problem is related to one of the standard Babylonian problems: Find x, y, if the product and difference are given. That is, solve the system

$$xy = c, \qquad x - y = b.$$

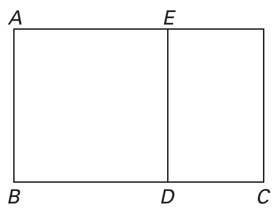


FIGURE 3.29
Data, Proposition 84

Euclid began by setting up the rectangle contained by the two straight lines AB, BC (Fig. 3.29). He then chose point D on BC so that $BD = AB$. Thus, $DC = b$ was the given straight line. He now had a given area, the rectangle ($= c$) applied to a given line b, exceeding by a square figure. He could then apply Proposition 59:

PROPOSITION 59 *If a given area be applied to a given straight line, exceeding by a figure given in form, the length and width of the excess are given.*

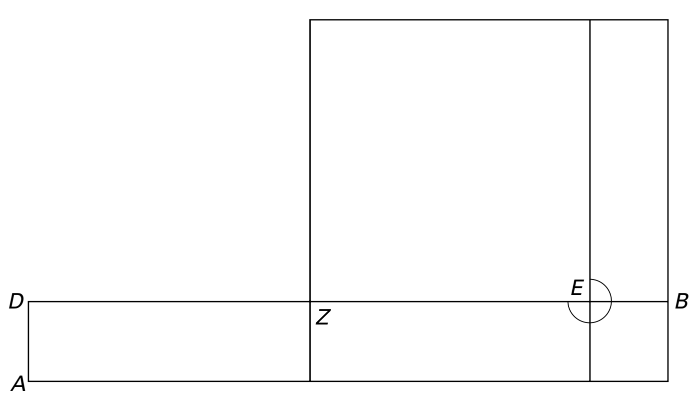
It is here that Euclid really solved the problem of Proposition 84, using a diagram similar to that of *Elements* VI–29 (Fig. 3.30). As there, he bisected the line $DE = b$ at Z, constructed the square on $ZE = b/2$, noted that the sum of that square and the original area (the rectangle $AB = c$) is equal to the square on $ZB = y + b/2$ (or $x - b/2$), and thereby showed how either of those quantities can be determined as the side of that square. Algebraically, this amounts to the standard Babylonian formula

$$y = \sqrt{\left(\frac{b}{2}\right)^2 + c} - \frac{b}{2}$$

$$x = \sqrt{\left(\frac{b}{2}\right)^2 + c} + \frac{b}{2}.$$

As before, Euclid dealt only with geometric figures and never actually wrote out a rule like the above. Nevertheless, given that the problem is in fact to find two lengths satisfying certain conditions, even its formulation is nearly identical to the Babylonian formulation. On the other hand, as in VI–29, the statement of the result enables one to deal with parallelograms as well as the rectangles discussed by the Babylonians. Euclid treated other similar geometric algebra

FIGURE 3.30
Data, Proposition 59



problems in the *Data*. Thus, in Propositions 85 and 58 he solved the geometric equivalent of the system

$$xy = c, \quad x + y = b,$$

while in Proposition 86 he solved the system

$$xy = a, \quad \frac{y^2 - b}{x^2} = \alpha.$$

Most probably, in this latter problem, Euclid was showing that if two hyperbolas each have their axes as the asymptotes of the other, then their points of intersection are determined.[9]

That Euclid would present a problem useful in the study of conic sections is not surprising, given that he is credited with a book on the subject. And, as we noted earlier, many of the propositions in Book II have application to that subject as well. Besides his work in conics, Euclid is also credited with works in such fields as spherical geometry, optics, and music. Thus, whoever Euclid was, it appears from the texts attributed to him that he saw himself as a compiler of the Greek mathematical tradition to his time. Certainly, this would be appropriate if he was the first mathematician called to the Museum at Alexandria. It would therefore have been his aim to demonstrate to his students not only the basic results known to that time but also some of the methods by which new problems could be approached. The two mathematicians in the third century BCE who most advanced the field of mathematics, Archimedes and Apollonius, probably received their earliest mathematical training from the students of Euclid, training that in fact enabled them to solve many problems left unsolved by Euclid and his predecessors.

EXERCISES

1. Prove Proposition I–5, that the base angles of an isosceles triangle are equal to one another.
2. Find a construction to bisect a given angle and prove that it is correct (Proposition I–9).
3. Prove Proposition I–15, that if two straight lines cut one another, they make the vertical angles equal to one another.
4. Construct a triangle out of three given straight lines and prove that your construction is correct. Note that it is necessary that two of the straight lines taken together in any manner should be greater than the remaining one (Proposition I–22).
5. On a given straight line at a point on it, construct an angle equal to a given angle and prove that your construction is correct (Proposition I–23).
6. Prove Proposition I–32, that the three interior angles of any triangle are equal to two right angles. Show that the proof depends on I–29 and therefore on postulate 5.
7. Solve the (modified) problem of Proposition I–44, to apply to a given straight line AB a rectangle equal to a given rectangle c. Use Figure 3.31, where $BEFG$ is the given rectangle, D is the intersection of the extension of the diagonal HB and the extension of the line FE, and $ABML$ is the rectangle to be constructed.

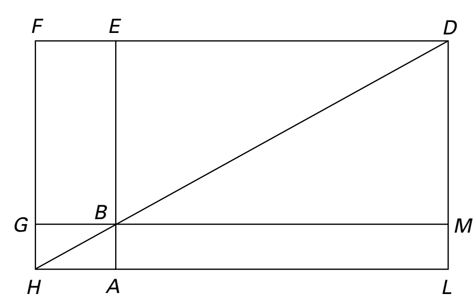


FIGURE 3.31

Elements, Proposition I–44