THE ROLES OF GEOMETRY AND ARITHMETIC IN THE DEVELOPMENT OF ALGEBRA: HISTORICAL REMARKS FROM A DIDACTIC PERSPECTIVE

LUIS RADFORD

In order to provide a brief overview of some of the historical affiliations between geometry and arithmetic in the emergence of algebra, we discuss some hypotheses on the origins of Diophantus' algebraic ideas, based on recent historical data. The first part deals with the concept of unknown and its links to two different currents of Babylonian mathematics (one arithmetical and the other geometric). The second part deals with the concepts of formula and variable. Our study suggests that the historical conceptual structure of our main modern elementary algebraic concepts, that of unknown and that of variable, are quite different. The historical discussion allows us to raise some questions concerning the role geometry and arithmetic could play in the teaching of basic concepts of algebra in junior high school.

1. THE ROOTS OF ALGEBRA: ARITHMETIC OR GEOMETRY?

1.1. *The Geometric Current*

The translation and interpretation of ancient Babylonian tablets, during the first half of this century, by Neugebauer (1935-1937), Neugebauer and Sachs (1945), and Thureau-Dangin (1938a) provide us with a wealth of knowledge on one of the earliest forms of mathematics practiced. Many of these tablets deal with numerical problems. In most cases, the problem-solving procedure is not completely explained and appears as a sequence of calculations. This makes it difficult to understand the way of thinking followed by the scribe in solving the problem. However, when interpreted in the wake of present day algebraic concepts and symbols, the calculations acquire some sense. This interpretation led the translators mentioned above, as well as some historians of mathematics (e.g., Boyer & Merzbach, 1991; Kline, 1972; van der Waerden, 1961, 1983), to claim that the Babylonians had developed a "Babylonian algebra." This algebra has been seen to be different from our modern elementary algebra, principally in its lack of symbolic representations.

The first problem of tablet BM 13901[1] is an example of this Babylonian algebra. Classically translated, the problem can be formulated as follows (cf., Thureau-Dangin, 1938a, p. 1, or van der Waerden, 1983, pp. 60-61) : "I have added the surface and the side of my square, and it is 3/4."

The classical interpretation of the solution is as follows:

> Take 1 to be the coefficient [of the side of the square]. Divide 1 into two parts. 1/2 x 1/2 = 1/4 you add to 3/4. 1 is the square of 1. You subtract 1/2, which you have multiplied by itself, and 1/2 is [the side of the square].

The classical interpretation "sees" the equation $x^2 + x = 3/4$ in the statement of the problem and "sees" the sequence of numerical operations leading to the solution as:

$$x = \sqrt{\left(\frac{1}{2}\right)^2 + \frac{3}{4}} - \frac{1}{2}$$

These calculations correspond to our own general formula for such a problem:

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

which yields the positive solution for the equation of type $x^2 + bx = c$.

Thus, according to the classical interpretation, Babylonian mathematicians would have known the general formula without being able to express it as such, because they lacked the symbols to do so.

The main argument supporting the idea of a Babylonian algebra is, thus, the possibility of translating the Babylonians' problems and calculations into modern arithmetic-algebraic symbolism. However it is an argument whose validity is not supported by historical evidence (see Unguru, 1975).

There is, however, a completely different interpretation of this type of problem. In fact, during the past years, Høyrup has studied the problems found in the Babylonian tablets as well as the terms used in them (Høyrup, 1987). Høyrup (1990) claims that:

> Old Babylonian "algebra" cannot have been arithmetical, that is, conceptualized as dealing with unknown numbers as organized by means of numerical operations. Instead it appears to have been organized on the basis of "naive," non-deductive geometry. (p. 211)

This non-deductive geometry, developed extensively in Høyrup (1985, 1986), consists of a "cut-and-paste geometry" in which the complicated arithmetical calculations resulting from the classical interpretation correspond to simple naive geometric transformations.

For example, Høyrup translates the problem shown above (BM 13901) as, "The surface and the square-line I have accumulated: 3/4."

His translation of the solution is as follows:

> 1 the projection you put down. The half of 1 you break, 1/2 and 1/2 you make span [a rectangle, here a square], 1/4 to 3/4 you append: 1, makes 1 equilateral. 1/2 which you made span you tear out inside 1: 1/2 the square line. (Høyrup, 1986, p. 450)

As one can see from the diagram which accompanies Høyrup's explanation,[2] the procedure consists of *projecting* a rectangle of base equal to 1 on the line-side of the square (Figure 1). The projected rectangle is cut into two rectangles, the base of each one being equal to 1/2. The rectangle on the right is then transferred to the bottom (Figure 2). Then a small square is added (Figure 3) in order to obtain a complete square from which the sought square-line can be found (Figure 4).

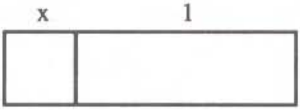


Figure 1.

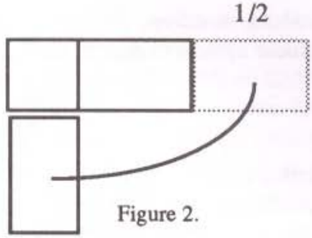


Figure 2.

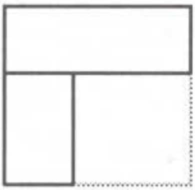


Figure 3.

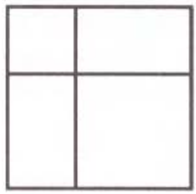


Figure 4.

This geometric interpretation is quite different from the classical interpretation which views such problems as problems dealing with numbers whose solutions are based on arithmetical (or algebraic) reasoning. Thus, while Neugebauer claims that "geometric concepts play only a secondary role" (Neugebauer, 1957, p. 65) in Babylonian algebra, Høyrup argues that *naïve geometry* is the basis on which such problems were posed and solved.

The new interpretation of the Babylonian algebra leads us to a re-evaluation of the role of arithmetic and geometry in the emergence of algebra.

We will discuss in Section 1.3 the influence of the "cut-and-paste geometry" in the emergence of algebra. Now we have to turn briefly to a Babylonian numeric current.

1.2. The *Numeric Current*

This new interpretation of Babylonian algebra, which concerns especially those problems whose translation into modern symbolism can be seen as second degree equations, does not preclude the existence of a numeric current in Babylonian algebra (see Høyrup, 1985, pp. 98-100). In fact, there are many problems, especially those concerning Babylonian commerce, which do not lend themselves to a geometric interpretation. This is also the case for certain first-degree problems (i.e., VAT 8520, No. 1) and the problems found on tablets VAT 8389 and VAT8391--both of which date back to the first Babylonian dynasty (circa 1900 B.C.).

The problems on tablets VAT 8389 and VAT 8391 concern grain production in two fields. If we designate α as the production per unit of area of the first field, X as the area of the first field, β as the production per unit of area of the second field, and Y as the area of the second field, most of the problems on the tablets can be formulated, using modern notation, as follows:

$$\alpha X \pm \beta Y = c; \qquad X \pm Y = \delta$$

(α, β, c, and δ having numerical values particular to each problem).

In modern notation, Problem 1 of tablet VAT 8389 corresponds to the following linear system of equations:[3]

$$\alpha X - \beta Y = c$$
$$X + Y = \delta$$

The calculations suggest that to find X and Y, the scribe first takes a *false solution*: in this case, $X_o = Y_o = \delta/2$ (satisfying the condition $X + Y = \delta$). He then calculates the production of each field (which he names "false grain"), which is, in modern notation, αX_o and βY_o.

Next, the scribe calculates the excess production of the first field over the second: $\alpha X_o - \beta Y_o$. Let c_o be this excess. He then calculates the production which is missing to satisfy the conditions of the problem. In our notation the production missing is $c - c_o$.

After that, the scribe calculates $\alpha + \beta$. This quantity is precisely the excess production of the first field over the second when we add one unit of area to the first field and we subtract one unit of area from the second field.

To compensate for the missing production $c - c_o$, we must make the excess $\alpha + \beta$ equal to $c - c_o$, a problem which can be solved through the tools of proportional quantities, a field of study in which the ancient pre-Greek mathematicians excelled. In fact, making the excess $\alpha + \beta$ equal to $c - c_o$ is achieved by multiplying $(\alpha + \beta)$ by the number of units of area to be added to X_o. The resulting quantity must then be equal to $c - c_o$.

The scribe knows that the number of units of area to be added to X, which we can designate as z, is obtained by multiplying the inverse of $(\alpha + \beta)$ by $(c - c_o)$. The number z thus obtained by the scribe is added to the first field and subtracted from the second field, thereby providing the real areas of the fields which are consequently $X_o + z$ and $Y_o - z$ respectively.

The procedure to solve this problem is clearly arithmetical and not geometric. The tablet allows us to see that it consists of an arithmetical method of *false position:* the scribe begins by assigning a numeric value, which is recognized as being false *a priori*, to the sought quantities (i.e., the areas of the fields). Using the false values and the data given in the problem's statement, he obtains new data. The new data (here the "false grain") can then be compensated for, to ultimately yield a correct solution. This method of *false position* is used to solve many of the problems on the Babylonian tablets (see, e.g., Thureau-Dangin, 1938b--tablets Str. 368, VAT 7535, and VAT 7532).

From a historical perspective, it is difficult to establish a link between the geometric and numeric currents in Babylonian algebra (see, however, Høyrup, 1990). It is also difficult to ascertain exactly what influence either of the currents may have had on the initial development of algebra. In fact, many of the most important early works containing basic algebraic concepts, such as Diophantus' *Arithmetica*, contain no explicit references to antecedent sources of inspiration. Nevertheless, we can trace certain elements in Diophantus' *Arithmetica* to the numeric and geometric currents of Babylonian and Egyptian mathematics.

1.3. *Diophantus' Arithmetica*

Diophantus' *Arithmetica* (circa 250 A.D.) is a collection of problems divided into 13 books, 3 of which remain lost. To clearly trace the links between Diophantus' algebra and some antecedent mathematical traditions, we must first recall that Diophantus, like Aristotle, conceived of number as being composed of discrete units. Moreover, Diophantus believed that numbers can be divided into "categories" or "classes," each category containing the numbers that share the same exponent: the first category being the squared numbers (designated as Δ^{Υ}), the second being the cubed numbers (designated as K^{Υ}), the third being the squared squares (designated as $\Delta^{\Upsilon}\Delta$) and the cubo-cubes (designated by $K^{\Upsilon}K$). The problem statements are written as *relations* between these categories. Hence, there are no particular numbers given in the problem statements. Following are two examples.

Book IV, Problem 6:

We wish to find two numbers, one square the other cubic, which comprise [i.e., that their product is] a square number. (Sesiano, 1982, p. 90)

Book I, Problem 27:

Find two numbers such that their sum and their product equal the given numbers. (Ver Eecke, 1959, p. 36--our translation).

After having classified numbers on the basis of their exponents at the beginning of Book I, Diophantus introduces one of the most important concepts to this discussion: the *arithme* (the "*number*") which is "an undetermined quantity of units" (Ver Eecke, 1959, p. 2). The arithme is introduced for the purpose of representing the unknown in a problem; it is a heuristic tool in the context of problem solving.[4]

Problem 27, Book I of the *Arithmetica*, referred to above, establishes a link between Diophantus' algebra and antecedent mathematical traditions. The solution begins as follows:

The square of half of the sum of the numbers we are seeking must exceed by one square the product of these numbers, which is figurative. (Ver Eecke, 1959, p. 36--our translation)

Thus, Diophantus begins by giving a condition which the numbers must fulfill, in order that the problem could be solved. The fact that this condition is "figurative" suggests that Diophantus is referring to a condition which can be visualized through a geometric representation (Ver Eecke, 1959, p. 36-37). Furthermore, Diophantus does not make his thought explicit, which suggests that he is referring to something well known to the reader (cf., Høyrup, 1985, p. 103). It is possible that Diophantus is actually referring to a cut-and-paste procedure (see Figure 5):

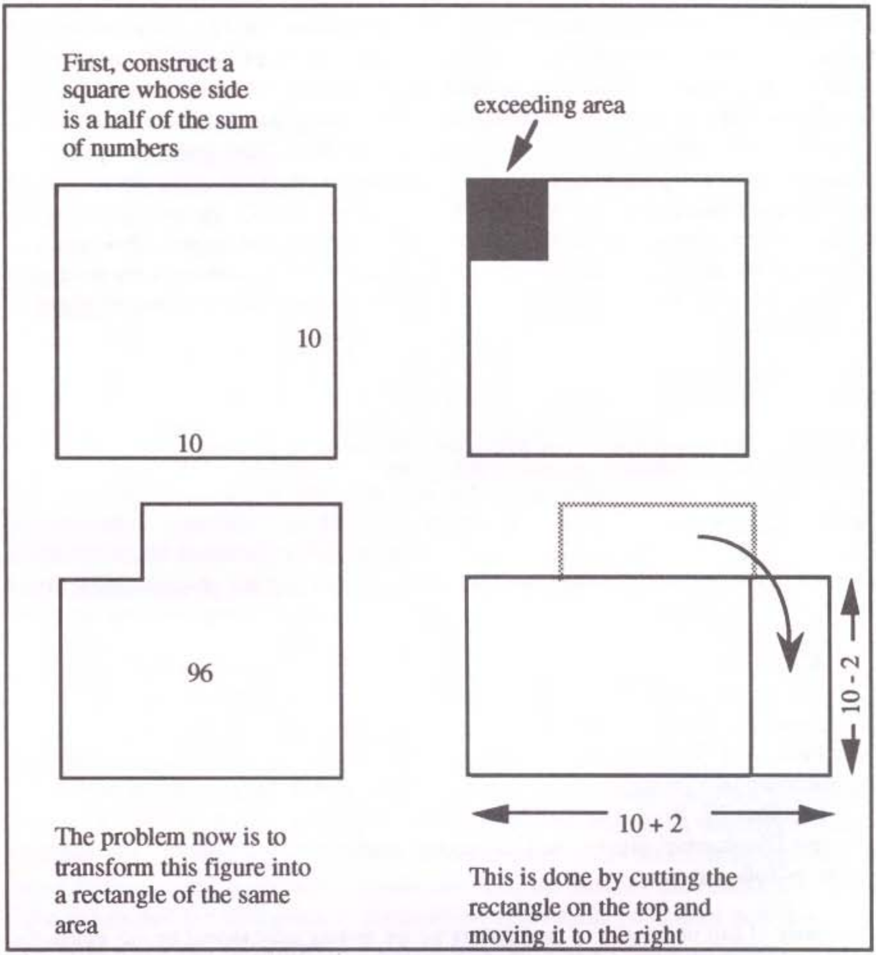


Figure 5.

For the problem to be solved numerically (using positive rational numbers, the only numbers considered by Diophantus), the exceeding area must be a square. As we see, the "figurative" condition mentioned by Diophantus seems to refer strongly to the above cut-and-paste procedure and suggests a link between Diophantus' methods and those of the Babylonian mathematicians belonging to the geometric current. To continue, Diophantus' solution to this problem is as follows: "Let the sum of the numbers equal 20 units, and their product equal 96 units." Diophantus then specifies the parameters of the problem:

Let the quantity in excess of the two numbers be 2 arithmes. Thus, because the sum of the two numbers is 20 units, which when divided by two yields two equal parts, each part will be half of the sum, or 10 units. Therefore, if we add to one part, and subtract from the other, half of the quantity in excess of the two numbers, that is 1 arithme, it follows that the sum will still equal 20 units and that the quantity in excess of the numbers will remain 2 arithmes. Thus, let the greater number be 1 arithme increased by 10 units, which is half of the sum of numbers, and consequently, the lower number is 10 units minus 1 arithme; it follows that the sum of these numbers is 20 units and that the quantity in excess of the numbers is 2 arithmes.

The product of the numbers must equal 96 units. This product is equal to 100 units minus 1 squared arithme; which we equate to 96 units, thereby causing the arithme to be 2 units. Consequently, the greater number will be 12 units and the lesser number will be 8 units. These numbers satisfy the initial conditions. (Ver Eecke, 1959, pp. 37-38--our translation)

To find the greater and lesser numbers, Diophantus used a method very similar to the one used in tablet VAT 8389 (a method that van der Waerden, 1983, pp. 62-63, called the "sum and difference"). In fact, both procedures take half of the sum of the numbers into account. They find a number that is added to and subtracted from half of the sum of the numbers. This resemblance of methods strongly suggests a link between Diophantus' methods and those of the Babylonian mathematicians belonging to the numeric current.

The similarity between Diophantus' method and the "false position" Babylonian method has been discussed by Gandz (1938) and by Thureau-Dangin (1938a, 1938b). Thureau-Dangin claims that the *false position* is in fact an *algebraic procedure,* where the unknown is *represented* by the number "1," for the Babylonians did not develop any symbolism allowing them to represent the unknown by a letter. However, this argument rather forces the old Babylonian thinking to fit into the modern algebraic thinking. On the other hand, Gandz states that Diophantus' algebraic method, shown in Book I, Problems 27-30, is a method which was known by the Babylonian mathematicians and which allowed them to find the solutions for the mixed second-degree equations. However, it supposes that ancient Babylonian mathematicians had at their disposal a complex symbolic language (see Gandz' calculations, pp. 416 ss.), but there is no historical evidence for this.

On the other hand, we can appreciate that Thureau-Dangin's and Gandz' main aim is to try to understand Babylonian mathematics. They go from Diophantus to Babylonian mathematics. We try to go in the other direction: from the Babylonians to Diophantus. We think that ancient mathematicians developed arithmetical methods, such as that of *false position*, in order to solve problems. Such a method is not based on algebraic ideas but on an idea of *proportionality*.

Yet another link between the arithmetical current and Diophantus' *Arithmetica* can be found in an Egyptian papyrus (known as Michigan papyrus 620) written before Diophantus (circa 100 A.D.) (Robbins, 1929). This papyrus contains a series of numerical problems that closely resemble many of the problems in Diophantus' Book I. Here is an example:

There are four numbers, the sum of which is 9900; let the second exceed the first by one-seventh of the first; let the third exceed the sum of the first two by 300, and let the fourth exceed the sum of the first three by 300; find the numbers.

This papyrus contains the symbol ς which Diophantus used in *Arithmetica*. Furthermore, the choice of the unknown in the papyrus is 1/7 of the first number--

not unlike Diophantus' choice of the unknown in his problems (a choice which is surely made to avoid calculations involving fractions: e.g., Book I, Problem 6).

To sum up our discussion, we can say that, from a historical point of view, the roots of Diophantus' algebra seem to be found in both: (a) a geometric Babylonian tradition (related to a cut-and-paste geometric current) and (b) a numeric Babylonian-Egyptian tradition (related to the false position method).[5] These mathematical traditions led to two kinds of different ancient "algebras":

1) On one hand, the "algebra" related to geometry, cultivated, in all likelihood, within the communities of surveyors who dealt with problems about geometrical figures. One of the paradigmatic problems of this current is to find the length or width of a rectangle satisfying certain conditions (e.g., the surface and the side equal to 3/4, as in BM 13901 seen above). In this algebra, the side of a figure can be seen as a side as well as a rectangle--the rectangle whose height (or *projection*) is equal to 1. Thus, the length of a side can also represent the area of the projected rectangle. This kind of algebra is underlined by the visual conservation of areas of figures submitted to the cut-and-paste procedures. Algebraic equality refers here to equalities between areas. Algebraic methods are essentially based on a sequence of geometric transformations, T_i, starting with the given figure, F_1 and ending with a square F_n of a known area:

$$F_1 \xrightarrow{T_1} F_2 \xrightarrow{T_2} \dots \xrightarrow{T_{n-1}} F_n$$

The unknowns of the problem (e.g., the lengths of the sides of a rectangle) are taken into account through the problem-solving procedure. However, the unknowns are not the object of calculation.[6]

2) On the other hand, the numeric tradition led to a "numeric algebra" that dealt with theoretical problems about numbers (such as those found in Diophantus' *Arithmetica*, but with some *riddles* also; e.g., to find the amount of apples divided between a certain number of people according to specific conditions). This kind of algebra is based on a different conceptualization from the geometric one. Here one does not have segments to represent the unknowns. Furthermore, there is not a "natural" name (like "side" or "height") to be used in speaking about the unknowns. A closer look at the problem-solving methods included in the *Arithmetica* shows that all the numeric algebra deals not with several unknowns but with a single unknown. Diophantus just called the unknown, the arithme, that is, the number (which should be understood as *the* number) for which we are looking. As in the case of the algebra of the geometric tradition, the numeric algebra supposes a hypothetical thinking: one reasons as if the number sought was already known. However, in contrast to the algebra of the geometric current, in the numeric algebra one calculates with and on the unknown. The calculation with/on the unknown makes it possible for the emergence of a new kind of calculation that is independent of the context and of the problem, a *formal calculus* (in the sense that it takes into consideration only the form--the ειδοσ, *eidos*--of the mathematical expressions). Within this new calculus, algebraic equality refers to equality between *species* (Radford, 1992),

something that we can translate into modern terms by *monoms*. The algebraic methods used to solve problems are based here on transformations of monoms. The goal of these transformations is to arrive at a monom equal to another monom (i.e., in modern terms, to an equation of the form $a \cdot x^n = b \cdot x^m$), and then to arrive at an equation of the form, a monom equal to a number (i.e., $c \cdot x^k = d$). As we can see, the heuristic behind the resolution of a problem in each algebra is different.

The mutual influence of these two kinds of algebras is difficult to detect in Babylonian mathematical thought. However, this influence is perceptible in Diophantus' *Arithmetica*, as we saw before in our example of Problem 27, Book I. Later on, these algebras seem to have followed separate paths. We can see cut-and-paste geometry reappearing in Arabic soil some centuries later, in the work of Al-Khwarizmi and in that of Abû Bakr. The algebras of numeric and geometric origin will meet together during the awakening of sciences in the late Latin Mediaeval Age, as a result of an intensive translation of Arabic and Greek mathematical works into Latin. Without a doubt, both algebras merged in the *Liber Abbaci* of Leonardo Pisano, in the beginning of the 13th century. Even though our modern elementary algebra looks more like the numerical algebra than the cut-and-paste one, we should be aware that the development of mediaeval algebra (and to some extent the algebra of the Renaissance) was organized according to some "types of equations" whose distinction was completely guided by cut-and-paste geometry.[7]

Concerning the ancient algebra of the numeric current, the historical records available today make it possible to reconstruct a scenario of the conceptual relationships between the false position methods and the Diophantine algebraic ones. We cannot discuss here such a scenario. Let us simply mention that the "jump" from arithmetic thinking to algebraic thinking seems to be located in a reinterpretation of the false position method based on the search for a shorter and direct method for solving problems according to which one no longer thinks in terms of false values but in terms of the unknown itself.

2. UNKNOWNS AND VARIABLES: TWO DIFFERENT CONCEPTUALIZATIONS

Our discussion thus far has centered on the concept of unknown. There is, however, an equally important concept which we have yet to consider: the variable. While the unknown is a number which does not vary, the variable designates a quantity whose value can change. A variable *varies* (see Schoenfeld & Arcavi, 1988, p. 421).

Where can we find the origin of variables in the history of algebra? As in the case of the unknown and in light of our current historical knowledge, it is difficult to accurately pinpoint the variable's "big-bang." In a certain sense, we find some traces of the concept of variable in some ancient Babylonian tablets. For instance, there are tablets for reciprocal numbers. But these tablets seem to stress a *relationship* between numbers rather than a variational property of a mathematical object. We can, however, trace certain elements of a more elaborate concept of variable in Diophantus' book entitled, *On Polygonal Numbers.*

In order to better understand the scope of Diophantus' conceptualization of variables, it is important to note that polygonal numbers emerge in a philosophical context of classification of numbers, which dates back to the era of the first

Pythagoreans.[8] An important book which includes a detailed treatment of polygonal numbers is Nicomachus' *Introduction to Arithmetic* (2nd century A.D.). In this work, Nicomachus tries to discover patterns between numbers, for example, that every square number is the sum of two consecutive triangular numbers (D'Ooge, 1926, p. 247).

However, the observed patterns are not proved (at least, not in the Euclidean sense; this is an altogether different type of mathematics in which philosophical considerations require no *deductive* proof).

A result, concerning triangular numbers, probably obtained from a few concrete numerical examples, is stated by Plutarch (who lived in Nicomachus' time) in the following terms: "Every triangular number taken eight times and then increased by 1 gives a square" (Heath, 1910/1964, p. 127). It is the generalization of this result to other polygonal numbers that interested Diophantus in his book, *On Polygonal Numbers.*

The book, of which the last part is lost, consists of four *deductively* connected propositions, concerning arithmetical progressions. The third, for instance, states that in an arithmetical progression "the sum of the largest and smallest [terms], multiplied by the quantity of numbers, form a number [equal to] twice the sum of the given numbers"[9] (Ver Eecke, 1959, p. 280--our translation). Using the first three propositions, Diophantus proves the fourth one. Using modern symbols, this proposition can be stated as:

$$S_n \text{ x } 8d+(d-2)^2=\left[(2n-1)\ d+2\right]^2$$

where n designates the side of the polygonal number, d designates the difference, and S_n designates the polygonal number.

Using the fact that $a = d + 2$ (cf., Footnote 8), the above proposition can also be written as:

$$S_n \text{ x } 8\ (a-2)+(a-4)^2=\left[(2n-1)\ (a-2)+2\right]^2$$

Diophantus uses this last proposition to obtain an *explicit formula* to calculate the polygonal number S_n when the "side," n, is known. This proposition also allows Diophantus to give a formula to calculate the "side," n, when the number S_n is known.

Translated into modern notation, the first of the formulas can be written as:

$$S_n=\frac{\left[(2n-1)\ (a-2)+2\right]^2-(a-4)^2}{8\ (a-2)}$$

(Take twice the side of the polygonal number; from this subtract one unit; multiply the result by the number of angles minus 2; then add 2 units. Take the square of the resulting number. From this, subtract the square of the number of angles minus 4. Divide the result by 8 times the number of angles minus 2 units. This gives us the polygonal number we are looking for.) (based on the translation of Ver Eecke, 1959, pp. 290-291)

It is important that we stress the conceptual nature of the numbers S_n, n, d, and a in the preceding propositions (or rather, their link to the concept of variable). To do this, we can refer to certain passages in the proof of proposition 3 mentioned above, that is: "the sum of the largest and smallest [terms], multiplied by the quantity of numbers, form a number [equal to] twice the sum of the given numbers."

The proof is divided into two cases: the first case is related to an even quantity of numbers and the second, an odd quantity of numbers. In the first case, Diophantus develops his reasoning taking into account only six numbers, represented by: α, β ,γ, δ, ε, ζ. (In the second case he represents five numbers.)[10] The fact that he limits himself to six or five numbers is a result of the symbolic system of representation he uses (a system which is in fact borrowed from Euclid; for an example, see *Elements*, Book IX, prop. 20); nevertheless, the generality of the proof (in the Greek sense) is not lessened by its limitation to six (or five) numbers.

The quantity of numbers (which we would call "*n*" in our modern symbolic system) is represented by a *segment.*[11] This is the segment $\eta\upsilon$ in Figure 6. In accord with Diophantus' concept of number, the segment $\eta\upsilon$ can be divided into its units: λ, μ, χ...

The "numbers in equal difference," that is, the numbers α, β, γ, δ, ε, ζ, are placed on the *segment.*

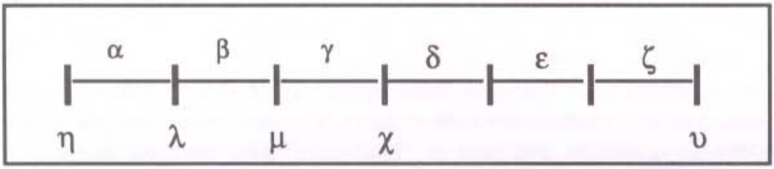


Figure 6. (From Tannery, 1893, Vol. II, p. 457)

This geometric representation allows Diophantus to organize the proof, which begins by noting (translated into modern notation) that:

$$\gamma - \alpha = \zeta - \delta$$

because of the equal difference between numbers. Then, Diophantus transforms this equality into:

$$\gamma + \delta = \zeta + \alpha$$

which he also writes as:

$$\gamma + \delta = (\zeta + \alpha) \times \eta\lambda$$

because $\eta\lambda$ is equal to the unit.

In the same way, Diophantus obtains:

$$\varepsilon + \beta = (\zeta + \alpha) \times \lambda\mu.$$

We do not need to go beyond this point in the proof. For our purposes, we only need to see that the quantity "*n*" of numbers and the number S_n, which is equal to $\alpha + \beta + \gamma + \delta + \varepsilon + \zeta$, are still not seen as *variable* numbers. In fact, *n* and S_n do not vary during the proof; "*n*," for instance, is a given number, chosen from the beginning.

Neither are the numbers n and S_n *empirical set values* in Diophantus' work, as is the case in Nicomachus' development of Polygonal Numbers Theory. Let us explain this idea. Patterns in Nicomachus' work state relations between numbers, but what supports the validity of patterns (in addition to philosophical premises) is the fact that when the numbers to which he refers in the patterns are replaced by concrete numbers, the resulting calculation is true. Thus, the numbers involved in Nicomachus' propositions are set values, or in order to stress their relationship with the concrete numbers, we can call them *empirical set values.* However, Diophantus never gives a numerical example. The propositions in Diophantus' theory of polygonal numbers are supported by a *deductive organization,* which breaks with Nicomachus' concrete-arithmetical treatment of numbers, and confers on numbers a different status. We can say that numbers (such as the numbers n, d, and S_n) become *abstract set values.*

We have said before that the numbers n, d, and S_n are not variables in the proof of a proposition given by Diophantus. Significantly, from the moment Diophantus states explicitly that one must find the polygonal number when the *side is given* (which arrives after all the deductive treatment of Proposition 4 is done), the numbers n and S_n are no longer *abstract set values;* they become *dynamic quantities* that can have different values depending on the values taken by the other quantity. They become variable mathematical objects.

This is, of course, a long way from the concept of variable that we find in the 18th century or by Oresme in the 14th century where the variable is especially linked to continuous quantities and seen as "intensities" associated to "qualities" (see Clagett, 1968). From the medieval context of the study of motion of bodies, variables will appear as flowing quantities, and the main problem will be that of describing the effect of the variation of one variable over another. This description will lead to the concept of *function.*

Diophantus is concerned about variables, not through the concept of function, but through the concept of formula. His concept of formula is not based on flowing continuous quantities, but on: (a) an explicit relationship between numbers that are seen as monads (i.e., unities) or fractional parts of monads and (b) an explicit sequence of calculations allowing one to determine a number given the identity of another number.

What is it, in Diophantus' work, that makes the emergence of the concept of variable possible? One reason may be that the problem is transformed from a *relational problem* to one involving *calculations*. This step can be achieved only in a "calculator's mind" (see Rey, 1948). The presence of the calculator's mind is obvious in the passages that we showed in the proof of Diophantus' Proposition 3: He uses the unity as a "real" number (and not only as the generating monad of the early Greek mathematicians). He performs calculations with the unity: this is why he can transform $\gamma + \delta = \zeta + \alpha$ into $\gamma + \delta = (\zeta + \alpha) \times \eta\lambda$.

This step requires--to pose the problem in the Greek context--that *Logistic* (which concerns the calculations of numbers, originally calculations with the material aspect of numbers, and later the way in which calculations are performed) be considered a theory of the same intellectual value as *Arithmetic* (which refers originally to the study of numbers in its deeper significance, in their essence, without any relation to the concrete world).[12] Arithmetic is concerned with propositions or theorems that state relations between numbers and/or properties of numbers. The step that joins *Logistic* and *Arithmetic* is not taken by Euclid, but is taken by Diophantus, in combining two intellectual inheritances: the Babylonian and Egyptian traditions of calculator mathematicians (as we have seen previously in Section 1) and the theoretical tradition of his predecessor in Alexandria, Euclid himself.

We now turn to the differences between unknowns and variables. The first difference is found in the context (more specifically, in the goal or intentionality) in which they appear. In fact, we can note that the main topic of *On Polygonal Numbers* is quite different from that of *Arithmetica*. In the former the goal is to establish *relationships* between numbers, while in the latter it is to solve word problems (i.e., to find the value of one or more unknowns). In *On Polygonal Numbers*, the relations between numbers are stated in terms of *propositions*; furthermore, they are organized in a deductive structure. Variables are derived from the passage from a *relational problem* to one dealing with *abstract set values calculations*. This is not the case in *Arithmetica*, where the problems refer to relations between "classes" or "categories" of numbers (squares, cubes, etc.) that constitute the Theory of Arithmetic.

The second difference between unknowns and variables is found in their representations. In the case of *On Polygonal Numbers*, the key concept is the *abstract set value* (which leads to that of variable). Abstract set value is represented geometrically or by letters, which are included in a geometric representation. In the case of *Arithmetica*, the key concept is the *unknown* (the arithme) which is not represented geometrically.

While both of these concepts deal with numbers, their conceptualizations seem to be entirely different.[13]

3. IMPLICATIONS FOR THE TEACHING OF ALGEBRA

What role could the preceding historical analysis play in the modern day teaching of mathematics? It seems to me that the historical construction of mathematical concepts can supply us with a better understanding of the ways in which our students construct their knowledge of mathematics. But aside from its contribution to our comprehension of certain phenomena arising in the psychology of mathematics, the epistemology of mathematics can also provide us with information that allows us to improve our teaching methods. Thus, we can ask ourselves certain questions:

1) Is it to our advantage to develop certain elements of "cut-and-paste geometry," as developed by Høyrup, for use in our classrooms in order to facilitate the acquisition of basic algebraic concepts? Can the teaching of "cut-and-paste procedures" awaken in students the analytical thinking that is required in learning algebra?
2) Is it to our advantage to introduce certain elements of the "false position" method prior to introducing the concept of the unknown in the solution of word problems?
3) Could we use proportional thinking as a useful link to algebraic thinking?
4) Can current teaching methods introduce some appropriate distinction between the concepts of unknown and variable, using the historical ideas seen here?

5) Is it reasonable to believe that the development in a deductive context of the concept of *abstract set value*, as developed by Diophantus, is a prerequisite to a deep learning of the concept of variable?

These questions deserve a great deal of reflection and discussion, which are beyond the scope of this chapter. Question 5, for instance, suggests a way of thinking about variables that contrasts with the usual way taken to introduce variables in junior high school, where the inductive context (based on the empirical set value concept of number) is preferred to a deductive one. Of course, from a behavioral point of view, the inductive tasks can be accomplished more quickly in the classroom than the deductive ones. But in terms of real understanding of what a variable is, our analysis can at least indicate some new didactic dimensions to be explored, in order to reach a deeper knowledge of the concept of variable.

Our analysis also suggests that symbolism does not capture completely the deep meaning of the concept of variable, as the teaching of mathematics seems to imply.

However, the results of this epistemological analysis should not be construed as normative for teaching. Rather, such an analysis is meant to give us a better understanding of the significance of mathematical knowledge and to provide us with a path, one of many, for its construction.

NOTES

1 The numbers in the following two translations are in base 10 (and not in base 60); as well, all fractions are represented using the modern symbol "/".

2 This diagram is implicitly referred to in the problem-solving procedure through the *meaning* of geometrical operations (see Høyrup, 1985, pp. 28-29).

3 The statement of this problem (based on Thureau-Dangin's French translation, 1938a, p. 103-105) is as follows: "By *bur*, I perceived 4 *kur*. By second *bur*, I perceived 3 *kur* of grain. One grain exceeds the other by 8`20 <*sila*>. I added my fields: 30` <SAR> What are my fields?" (NB: *bur* is a measure of surface; *kur* and *sila* are measures of capacity. A complete commented solution of this problem can be found in Radford, 1995a).

4 The arithme is designated as the letter ς, probably, as Heath (1910/1964) suggested, because it is the last letter in the Greek word arithme (arithmos, αριθμος). It is important to note that the symbolism used to designate the arithme and its powers (square, cube, etc.) leads us to the first symbolic algebraic language ever known (see Radford, 1992).

5 There are other problems in Diophantus' *Arithmetica* whose problem-solving procedures recall the false-position method: for example, Book "IV" problem 8 (Ver Eecke, 1959, pp. 119-120); Book "IV" problem 31 (Ver Eecke, 1959, pp. 155-157). These problems are discussed in Katz (1993, pp. 170-172).

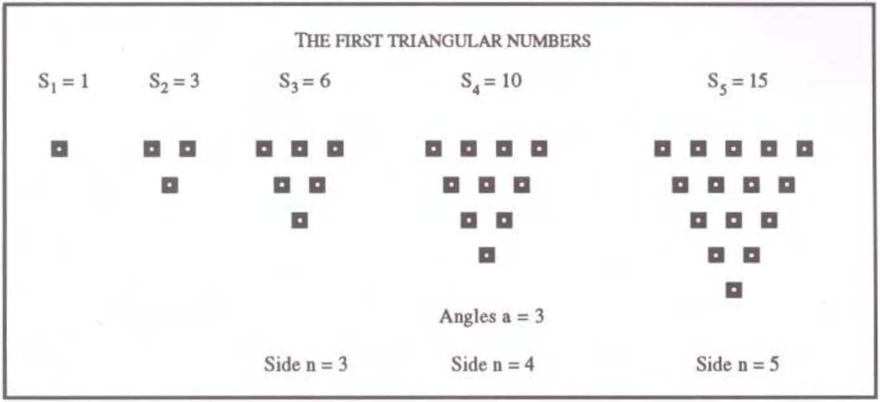
6 For instance, if we refer to the first problem of tablet BM 13901, shown earlier in this chapter, we can see that the unknown is found by displacements of figures. The unknown is not *really* involved in calculations, which are performed or executed with known quantities (Radford, 1995a). Although sometimes the scribe takes a fraction of the unknown or the unknowns in order to cut a figure (for some examples, see Høyrup 1986, pp. 449-455), it does not constitute a truly extended or generalized calculation ***with*** or ***between*** unknowns. The limits of such a geometric calculus can be better circumscribed if it is compared to the algebraic calculus which Diophantus develops at the beginning of his *Arithmetica* (see Ver Eecke, 1959, pp. 3-9).

7 See Radford (1995b).

8 Remember that polygonal numbers are obtained as the sum of the first numbers in an arithmetical progression with difference d and first term $a_1 = 1$. If we designate, in modern notation, the arithmetical progression as $1, 1 + d, 1 + 2d, \ldots$, then the polygonal numbers are:

$$S_1 = 1,\ S_2 = 1+(1+d),\ S_3 = 1+(1+d)+(1+2d), \ldots, \text{etc.}$$

If $d = 1$, we obtain the triangular numbers, if $d = 2$, we obtain the square numbers, if $d = 3$, we obtain the pentagonal numbers, etc. The first triangular numbers are: 1, 3, 6, 10, ... Note that the angles a of a polygonal number are obtained by adding 2 to its generating difference d, that is $a = d + 2$. Thus, the *angles* of any triangular number are $a = 1 + 2 = 3$. The *side* of a specific polygonal number $S_n = a_1 + a_2 + ... + a_n$ is n. Thus the side of the triangular number 6 is 3.



9 In modern notation, this proposition can be stated as: $\left(a_n + a_1\right)n = 2S_n$.
10 A discrete representation of the numbers (i.e., by small circles or other objects, as used by Nicomachus), would not have suited this proof. In fact, in order to keep the numbers unspecified, new "abstract" representations are required. The choice of letters as representations satisfies this requirement.
11 It should be noted that geometrical segments are used here in a completely different way from that of Babylonian mathematics! (see Section 1.1).
12 The next passage from Olympiodorus' *Scholia* to Plato's *Gorgias* explains the initial difference between *logistic* and *arithmetic*: "It must be understood that the following difference exists: arithmetic concerns itself with the kinds of numbers; logistic, on the other hand, with their material" (cited by Klein, 1968, p. 13).
13 The distinction between variables and unknowns is not stated explicitly by Diophantus. This distinction is explicitly made in the 18th century by Leonard Euler, who sees unknowns as objects belonging to "ordinary analysis" (i.e., algebra) while variables are seen as objects belonging to "new analysis" (i.e., infinitesimal analysis) (see Sierpinska, 1992, p. 37).