Philosophia Mathematica (III) **14** (2006) 287–317 doi:10.1093/philmat/nkj017
Advance Access publication January 23, 2006

The History of Algebra and the Development of the Form of its Language†

LADISLAV KVASZ*

This paper offers an epistemological reconstruction of the historical development of algebra from al-Khwārizmī, Cardano, and Descartes to Euler, Lagrange, and Galois. In the reconstruction it interprets the algebraic formulas as a symbolic language and analyzes the changes of this language in the course of history. It turns out that the most fundamental epistemological changes in the development of algebra can be interpreted as changes of the pictorial form (in the sense of Wittgenstein's *Tractatus*) of the symbolic language of algebra. Thus the paper develops further the method of reconstruction which the author introduced for the analysis of the development of geometry.

While geometry represents a classical topic in the philosophy of mathematics—it suffices to mention Plato, Pascal, Spinoza, Kant, or Husserl—the philosophical questions posed by algebra have remained on the periphery of interest. Apart from the books of Jacob Klein [1934] and Jules Vuillemin [1962], little attention has been paid to algebra in the philosophical literature. This is a pity, because a philosophical interpretation of the insolubility of the equations of the fifth degree, for example, could shed light on the relation between language and reality. The aim of the present paper is to describe the fundamental changes in the development of algebraic thought and to reflect on their philosophical significance. To this end I will use an approach that I introduced in the analysis of the development of geometry [Kvasz, 1998]. In particular I will divide the development of algebra into distinct periods using the terms from this earlier study, such as '*perspectivist form of language*' or '*projective form of language*'.

The notions of perspective or projection are doubtless geometrical notions. Nevertheless, I use them in an epistemological sense to charac-

† I would like to thank Donald Gillies, Táňa Jajcayová, Eberhard Knobloch, Paul Rusnock, and Pavol Zlatoš for their comments and criticism of previous versions of this paper. I would like also to express my gratitude to the Alexander von Humboldt Foundation for a scholarship at the Technical University in Berlin, which made it possible to write this paper.

* Department of Algebra, Geometry and Mathematics Education, Comenius University, 84248 Bratislava, Slovakia. kvasz@fmph.uniba.sk

terize a particular form of language. For the perspectivist form, for instance, it is typical to consider expressions of language to be pictures of reality seen from a particular point of view, while the projective form introduces as a fundamental innovation the representation of a representation. In geometry the transition from the perspectivist to the projective form is obvious. I would like to show that from the epistemological point of view a very similar transition also occurred in the development of algebra. At the beginning mathematicians accepted as a solution of an equation only a unique positive value of the unknown that fulfilled all the conditions stated in the formulation of the mathematical problem. The unknown was thus seen as a 'picture' of a quantity of the real world, precisely as in the perspectivist representation. Later, when mathematicians introduced substitutions (*i.e.,* representations of representations) the view arose that all roots of an equation are to be accepted as solutions. This meant that not only the '*true*', that is, positive solutions, those with a direct relation to reality, were accepted but also (to use Descartes's term) the '*false*' and even the '*imaginary*' ones. The solution thus ceased to be a 'picture' of the extralinguistic reality and became simply an element of a structure. Substitutions in many respects resemble the projections of projective geometry. Both introduce a representation of a representation. And just as the introduction of projections into geometry led Desargues to complement the plane by infinitely remote points in order to make projections one-to-one mappings, so too in algebra it turned out to be necessary to complement the number system by negative and complex numbers in order to ensure appropriate closure properties. Thus we see an analogy between projections of a geometric figure and transformations of an algebraic equation.

When we transfer the notion of the form of language from geometry to algebra we discover that objects like 0 or 1 play a role in algebra similar to that played in geometry by notions like center of projection or horizon. The role played by the form of language is to integrate the epistemic subject of the language into the universe of the language. In geometry, where the universe of the language lies before us and the epistemic subject has the form of a point of view, the position of this subject in the universe is fixed by means of the horizon. The horizon indicates the horizontal plane of our sight. In algebra, when compared with geometry, the situation is rather the opposite. The universe of algebra is created through the reification of the activities of the epistemic subject. The numbers 3 or $\sqrt{5}$ are symbolically reified acts of counting or of extracting the root. Thus the universe of the language of algebra is not a region opened to our gaze where our only task is to fix our position. On the contrary, the universe of algebra is something that was originally part of ourselves, namely, of our actions, and we only distance ourselves from its objects step by step as we represent the actions like counting, adding, subtracting, *etc.* by symbols of numbers, addition, or subtraction. In this way the reification of arithmetical operations like addition or root extraction brought the deeper structure of symmetries to the fore, which was in its turn reified in the form of a group.

The universe of algebra is thus not an external world open to our gaze. The universe of algebra arose in the process of gradual reification of linguistic descriptions of our activities. It is no accident that algebra was created by the Arabs rather than the Greeks. In our reconstruction of the development of algebra we will consider the history of algebra in western civilization. We will follow the history of western civilization, a civilization of the 'geometric spirit', a civilization for which to understand means to acquire an insight, and we will reconstruct how this civilization integrated algebra, a discipline the universe of which is not open to our gaze, a discipline that is based on the motoric experience, on manipulations, on actions. In contrast to the ideal of the Greek theoretician, who does not intervene in the events of the world and only follows them with his look, an algebraist always acts, performs calculations, solves equations, transforms expressions. In algebra, insights are possible only after an action, only after some lucky trick brought us to the sought solution. But besides such positive experiences, when we find the solution, in algebra very often we have to deal with a negative experience, with the experience that the transformation did not lead to the desired result. Such negative experiences also require conceptualization; they too lead to the desire to understand why the actions did not succeed. One of the central threads in the development in algebra was motivated by the wish to understand the systematic failure of all attempts to solve the quintic equation.[1] Why is it that despite the attempts of the best mathematicians during the period of three centuries nobody managed to solve such simple equations as $x^5 - 6x + 3 = 0$?

In situations of this kind, the problem is not to arrive at an insight, because at the beginning there is nothing on which our sight could rest. There is only a long line of failures, heaps of paper filling the wastebaskets of leading mathematicians—all this in thousands of variations. The drama of the development of algebra lies in the slow process by which mathematicians finally succeeded in determining the outlines of the invisible barrier blocking all attempts to solve the quintic equation. This barrier is called the alternating group of five elements. The mathematicians who succeeded in determining its outlines were Paolo Ruffini, Niels Henrik Abel, and Evariste Galois.

[1] A proof of the insolubility of the quintic equations can be found for example in [Stewart, 1973, p. 134].

0. The Prehistory of Algebra, or Why the Greeks Did Not Create Algebra

When we say that the Greeks did not develop algebra,[2] this does not mean that they were not able to solve mathematical problems that are now considered algebraic. Long before the Greeks, the ancient Babylonians achieved mastery in solving problems equivalent to what we today call quadratic equations. Nevertheless, they solved these problems without the use of symbolic manipulations, relying solely on the art of reckoning. The Greeks, on the other hand, excluded reckoning from mathematics, and reduced the problems that we consider algebraic to geometry. So they lost any contact with the calculative way of thinking predominant in ancient Babylonian mathematics. For Euclid the problem that we today understand as a quadratic equation $x^2 + bx = C$, and solve using the well-known formula, was a geometric problem. The task was to find a segment x such that a square having it as a side, together with a rectangle having this segment as one side and b as the other, together have an area equal to the area of a given figure C. Euclid (VI, problem 28) shows us how to construct the segment x. But even though Euclid was able to provide nice constructions for quadratic equations, the geometric approach to such problems has a fundamental limitation. When we interpret x^2 as the area of a square, then x^3 is the volume of a cube, and we must stop there, because the higher powers have no counterparts in three-dimensional space. This was why the Greeks, confined to geometry, were able to capture only a small fragment of the world of algebraic equations, a fragment that may have been too limited to give rise to an independent mathematical discipline. This may be why the world of algebra remained hidden from the Greeks.

When we examine Arabic algebra, we find that from the point of view of technical finesse it does not surpass the level of Euclid's *Elements*. On the contrary, in comparison with the complicated pattern of argumentation in Euclid and the sophisticated geometric constructions, Arabic algebra is in many respects simpler. Thus the reason why the Greeks did not invent algebra could not have been an inability to think abstractly. Rather, we get the impression that there was some obstacle preventing the Greeks from entering the universe of algebra. This universe was first entered by Islamic mathematicians. Of course, in mathematics the Arabs were pupils of the ancient Greeks. From the Greeks they learned what proofs, axioms, and definitions are. But their culture was different. In its center stood a

[2] The explanation of why the Greeks did not develop algebra can be only hypothetical. I base it on a strong dominance of the visual aspect of Greek geometry, and a nearly complete ignorance of the calculative aspect of arithmetic. We can obtain a deeper insight into this question if we look at it in connection with other aspects of early modern science, which had no parallel in ancient science (see [Kvasz, 2004]).

religion that refused to use the metaphor of sight in approaching the transcendent. In Arabic culture the close link between thought and visual experience vanished. This link was the core of the Greek *episteme*. The basic metaphor behind the Greek notion of theoretical knowledge is a *view from a distance*. In order to get an insight into a problem, we have to step back, we have to create a distance, we have to free ourselves from everything that ties us to it and which could therefore disturb the impartiality of our view. Only from a distance can we see the truth (ἀληθεια—that which is not hidden from our gaze). This tendency is clearly visible in Euclid's *Elements*, where the constructions are strictly separated from the proofs. During a proof we must not *do* anything; we have only to *look* in order to discern the truth.

From this it is clear why algebra, in which the focus lies on manipulations with formulas, was alien to the Greek culture. Algebraic manipulations are based not on theoretical insight but on skill in calculations. The aim is not to envision the result but rather to acquire a sense for the tricks that can lead to it, a sensitivity to the possibilities opened by the language of algebra. These possibilities are never actualized, they are not open to a gaze. In the algebraic language only the expression we are actually executing is actualized. Of course, we remember many previous calculations and manipulations, and we see the chain of calculations that we actually perform against the background of this past experience. We sense the analogies and similarities; we recall stratagems that can lead us through the thicket of possible steps to the result. Algebra is always about a process. When Arabic algebra came through Spain to Europe, a dialogue started between the spirit of western mathematics and this fundamentally different, but equally deep, spirit of algebra. This dialogue was dominated by attempts to visualize, the endeavor to get all the tricks and manipulations into the visual field, to acquire an insight into them. Nevertheless, whenever the western spirit succeeded in visualizing one layer of algebraic thought, another deeper layer emerged beneath it. In our reconstruction we will try to show how beneath the rules of algebra algebraic formulas emerged, then beneath algebraic formulas forms appeared, beneath forms fields, beneath fields groups. The history of algebra is a history of a gradual reification of activities, transforming operations into objects.

1. The Perspectivist Form of the Language of Algebra: The Solution of an Equation as a Rule (from Al-Khwārizmī to Cardano)

Abū Abdallāh Muḥammad ibn Mūsā al-Khwārizmī (780–850) is the author of the *Short book of al-jabr and al-muqābala*, a treatise on solving 'equations'. The word *al-jabr* (algebra) in the title of the book came in time to be used as a name for the whole discipline dealing with 'equations'. We cannot speak about equations in the modern sense, because the book of al-Khwārizmī, remarkably, makes no use of symbols, and even numbers are expressed verbally. For the powers of the unknown the book uses special terms: for x it uses *shai*' (thing), for x^2 *māl* (property), for x^3 *ka'b* (cube), for x^4 *māl māl*, for x^5 *ka'b māl*, *etc*. In translations of the work the Arabic names for the powers of the unknown were replaced by their Latin equivalents; thus *res* stood for *shai'*, *census* for *māl*, and *cubus* for *ka'b*. In Italy the word *cosa* replaced the Latin *res*, and so during the fifteenth and sixteenth centuries algebra was usually named *regula della cosa*—the rule of the thing. Algebra was understood as a set of rules for manipulating the thing (*i.e.*, the unknown), which enable us to find the solutions of particular 'equations'.

Before attempting to solve a particular 'equation', al-Khwārizmī first rewrote it in a form where only positive coefficients appeared and the coefficient of the leading term (term with the highest power of the unknown) was 1. In order to achieve this standard form, he made use of three operations: *al-jabr*—if on one side of the 'equation' there are members that have to be taken away, the corresponding amount is added to both sides; *al-muqābala*—if the same power appears on both sides, the smaller member on the one side is subtracted from the greater one on the other side; and *al-radd*—if the coefficient of the highest power is different from 1, the whole 'equation' is divided by it. We write the term 'equation' in quotation marks because, strictly speaking, al-Khwārizmī did not write any equations. Rather, he transformed relations among quantities, everything being stated in sentences of ordinary language, enriched by few technical terms.

We can illustrate his approach with an example. Consider the equation $x^2 + 10x = 39$, which he expressed in the form: '*Property and ten things equals thirty-nine*'. His solution reads as follows: '*Take the half of the number of the things, that is five, and multiply it by itself, you obtain twenty-five. Add this to thirty-nine, you get sixty-four. Take the square root, or eight, and subtract from it one half of the number of things, which is five. The result, three, is the thing*.' This approach is close to the Babylonian tradition. It is a set of specific instructions telling us how to find the solution. Nevertheless, there is a substantial difference. In contrast to the Babylonian mathematicians al-Khwārizmī has *the notion of the unknown* (shai') and therefore his instruction '*take the half of the number of the things, multiply it by itself, add this to thirty-nine, take the square root, and subtract from it one half of the number of things*' is a universal procedure, which can be applied to any quadratic equation of that particular form. Thus, while the Babylonian mathematicians only made their calculations with concrete numbers, al-Khwārizmī is able to grasp the procedure of solution in its entire universality. When he uses concrete numerical values for the coefficients, he does so only for the purpose of illustration. With the help of the notions *shai'*, *māl*, and *ka'b* he is able to grasp the universal procedure, and in taking this step he became the founder of algebra.

In the twelfth century the works of al-Khwārizmī were translated into Latin. The custom of formulating the solution of an 'equation' in the form of a verbal rule persisted until the sixteenth century. The first result of western mathematics that surpassed the achievements of the Ancients was formulated in this way. This was the solution of the cubic equation, and was published in 1545 in the famous *Ars Magna sive de Regulis Algebracis* by Girolamo Cardano (1501–1576). Cardano formulated the equation of the third degree in the form: '*De cubo & rebus aequalibus numero*'. The solution is given in the form of a rule: '*Cube one-third of the number of things; add to it the square of one-half of the number; and take the square root of the whole. You will duplicate this, and to one of the two you add one-half the number you have already squared and from the other you subtract one-half the same. You will then have* binomium *and its* apotome.[3] *Then subtracting the cube root of the apotome from the cube root of the binomium, that which is left is the thing.'*

In order to *see* what Cardano was *doing*, we present the equation in modern form $x^3 + bx = c$ and we express its solution with the help of modern symbolism:

$$x = \sqrt[3]{\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \left(\frac{b}{3}\right)^3}} - \sqrt[3]{-\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \left(\frac{b}{3}\right)^3}}.$$

Of course, Cardano never wrote such a formula. In his times there were no formulas at all. On the surface algebra was still *regula della cosa*, a system of verbal rules used to find the thing.

2. The Projective Form of the Language of Algebra: The Solution of an Equation as a Formula (from Regiomontanus to Descartes)

In the previous section I presented Cardano's rule for the solution of the cubic equation. Even if the rule itself did not deviate from the framework of al-Khwārizmī's approach to algebra, it is not clear how was it possible to discover something so complicated as this rule. In order to understand this, we have to go back a century before Cardano and describe the first

[3] Cardano called $\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \left(\frac{b}{3}\right)^3}$ a *binomium* and $-\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \left(\frac{b}{3}\right)^3}$ an *apotome*. Of course, Cardano did not write formulas, and what we here presented is a formal transcription of his verbal descriptions in modern algebraic symbolism. But despite this difference, these termini of Cardano show the necessity of referring to specific algebraic entities as if they were objects, and so document an early stage of the process of reification that led to the creation of algebraic symbolism.

stage of the reification of the language of algebra, the stage connected with the creation of algebraic symbolism. After western civilization had absorbed Arabic algebra, a tendency arose to turn algebraic operations into symbols and in this way to make them visible. This process was slow, lasting nearly two centuries, and most likely those responsible for it were not fully aware of it. We will present only some of the most important innovations. Regiomontanus (1436–1476) introduced the symbolic representation for root extraction. He denoted the operation of root extraction by *R*, stemming from the Latin *radix*. Thus for instance he expressed the third root of eight in the form *R cubica de 8*. In this way he represented the *operation* of root extraction by the *expression* of the root itself, that is by the result of the operation. Michael Stifel (1487–1567) replaced the capital *R* by a small *r*, so that instead of *R cubica de 8* he wrote $\sqrt{c}8$. He introduced the convention of writing the upper bar of the letter *r* a bit longer. Stifel placed the first letter of the word *cubica* below this prolonged bar, so that everybody would know that it was the cube root. The number placed after this sign is the one whose root is to be extracted. This convention is quite similar to that used by Regiomontanus, differing only in that it is a bit shorter, using only the first letter of *cubica* instead of the whole word. Nevertheless, this small change opened the door to our modern convention, which was introduced by Descartes (1594–1650). Descartes replaced Stifel's letter *c* by the upper index, and placed the number itself below the bar of the letter *r*, thus writing the third root of eight as $\sqrt[3]{8}$.

Another very important development took place in connection with the representation of the unknown. The Arabic terms of *shai'*, *māl,* and *ka'b* were translated as *res*, *zensus*, and *cubus*. Perhaps because the latter two are rather long, instead of writing the whole words mathematicians started to use only their first letters, thus *r* for *res*, *z* for *zensus*, and *c* for *cubus*. Just like the Arab algebraists, the *Cossists* (as the practitioners of this new algebra were called) did not stop with the third power of the unknown, introducing higher powers, such as *zz* (*zenso di zensi*), *zc* (*zenso di cubo*), *etc*., and developing simple rules for calculating with such expressions. Through such gradual processes symbols for the algebraic operations were introduced and an algebraic symbolism appeared. The operation of addition was represented by the symbol $+$, the operation of root extraction by $\sqrt{\ }$, and gradually a whole layer of operations was reified, *acts were turned into objects*. This process was slow, and at the beginning it was only little more than the replacing of words by letters for the sake of brevity. Nevertheless, when the new symbols accumulated in sufficient quantity, they made possible a radical change in algebraic thought—the solution of the cubic equation. As we saw, Cardano formulated his result in traditional form, as a verbal rule. Nevertheless, its discovery was made possible by the new symbolism. We will present a reconstruction of this discovery, presenting it in modern symbolism for the sake of comprehensibility (see [Scholz, 1990]).

Let us take a cubic equation

$$x^3 + bx = c,$$

i.e., Cardano's '*De cubo et rebus equalibus numero*'.[4] The decisive step in the solution of this equation is the assumption that the result will have *the form of* the difference of two cube roots. When we make this assumption, everything becomes simple. Let

$$x = \sqrt[3]{u} - \sqrt[3]{v}. \tag{1}$$

Raising this expression to the third power and then comparing it with the equation we obtain the following relations between the unknown quantities u and v, and the coefficients b and c:

$$b = 3\sqrt[3]{uv}, \quad c = u - v. \tag{2}$$

When we isolate v from the second equation, and substitute the resulting expression into the first one, we obtain a quadratic equation

$$u^2 - uc - \left(\frac{b}{3}\right)^3 = 0. \tag{3}$$

The root of this equation can be found by the help of the formula for quadratic equations as

$$u = \frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \left(\frac{b}{3}\right)^3}. \tag{4}$$

The value of the unknown v can be now determined from the second equation of (2). Knowing u and v we can find the solution of the original problem from (1) in the form

$$x = \sqrt[3]{u} - \sqrt[3]{v} = \sqrt[3]{\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \left(\frac{b}{3}\right)^3}} - \sqrt[3]{-\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 + \left(\frac{b}{3}\right)^3}}.$$

[4] The two remaining cases of cubic equations, the $x^3 = bx + c$ (*i.e.*, *De cubo equalibus rebus et numero*) and $x^3 + c = bx$ (*De cubo et numero equalibus rebus*) can be solved in a similar way. Equations of the form $x^3 + bx + c = 0$ were not considered, because all numbers occurring in an equation were taken to be positive so that their sum cannot be zero.

In this derivation the advantage of the algebraic symbolism, *i.e.*, of the reification of the language of the Arabic algebra, is clearly visible. Right at the beginning we assumed that the result would have *the form of* the difference of two cubic roots. In Arabic algebra there were no expressions, there were only rules. And *a rule has no form*, because it cannot be perceived. We can only listen to it, and then perform all the steps precisely as the rule instructs us. Only when we represent the steps of the rule by symbols does the sequence of calculations appear before our eyes; only then are we able to perceive its form. *The rule is thus transformed into a formula*.

Step (3) is also noteworthy. We obtained an equation for u. But what is this u? It is not a thing determined by the original equation. It does not stand for anything real. If we were to compare the equation to a picture we would see that here we have a *representation of a representation*, a picture inside a picture. From the technical point of view it is the decisive step in the whole process of solution, because instead of a cubic equation for the original unknown x, we obtain a quadratic equation for this auxiliary unknown, which we already know how to solve. From the epistemological point of view it is a fundamental shift. In the context of the original equation the unknown x has a real denotation: we know what it represents, what it stands for. But now it is replaced by some u, about which we know nothing. Its meaning is determined only through the equation (1). Thus the unknown u has no direct reference. Its *reference is given only indirectly*, through the reference of x.

Nevertheless, it is important to realize that new possibilities are opened up when one reifies one level of the algebraic language by turning the rules into formulas (*i.e.*, when the rule 'square the thing and add to it five things' becomes simply '$x^2 + 5x$'). Al-Khwārizmī knew only three algebraic operations—*al-iabr*, *al-muqābala*, and *al-radd*—but not *substitution*. His transformations did not make it possible to change the 'form' of the algebraic formulas. In this respect al-Khwārizmī's algebra resembles Euclidean geometry, which also studied only form-preserving transformations of geometric figures (parallel translations, rotations, and uniform changes of scale). Central projections, which can change the form of a figure (*e.g.*, it can transform a circle into a hyperbola), were studied in projective geometry. I would like to stress a fundamental analogy between central projections and substitutions. On the basis of this analogy, I suggest that we call this stage in the development of algebra, which is characterized by the emergence of substitutions, as the stage of the *projective form of the language of algebra*. The substitutions do not simply shift a term as a whole from one side of the equation to the other but rather decompose them and then rearrange them in a new way. For instance, the substitution $x = \sqrt[3]{u} - \sqrt[3]{v}$ decomposed the unknown x into two parts, rearranged these parts, and then put them together again. Thus it seems that the transition to the projective form of language *shifts the ontological foundations* one level deeper. Algebra as *regula della cosa* understood the unknown as a 'thing' that we can 'take in our hands and move to some other place', but the 'form' of this thing remained unchanged. In the projective form of language the 'thing' itself is transformed. It is, for instance, decomposed into two parts which can be treated separately.

Another important change introduced by the use of substitutions concerns what counts as the solution of an algebraic equation. Formerly, in the framework of algebra understood as *regula della cosa*, mathematicians accepted only positive solutions, because the number of the things (*cosa*) cannot be negative. Thus Cardano accepted only a positive root in his solution to the cubic equation. If the unknown represents some real quantity, some number of real things, it cannot be less than nothing. But as soon as we start using auxiliary equations, the unknowns of which refer only indirectly due to substitutions, it can happen that the positive solution of the original equation corresponds to a negative root of the auxiliary equation. Therefore at least in the auxiliary equations we have to take into account *the negative as well as the positive roots*. In this context Cardano distinguished between the 'false' and the 'true' solutions. The notion of an equation is thus slowly freed from its dependence on direct reference. To add the 'false' to the 'true' solutions was also necessary in order to turn the transformations of the formulas into equivalent transformations. Thus as a further aspect of the projective form of language we can mention the extension of the language by expressions with no direct denotation (infinitely remote points, negative roots), which make smooth functioning of the language possible.

The most important shortcoming of the previous algebraic notation was that it used different letters (r, z, c, ...) to represent the powers of the same quantity. Thus, for instance, if r is 7, then z must be 49, but this dependence is not indicated by the symbolism. When substitutions are used, such a convention becomes unwieldy, because whenever we make a substitution for r, we must also make the appropriate substitution for z. Further, in a substitution we have to do with at least two unknowns, the old one and the new one. To represent both with the same letter r would create an ambiguity. Another shortcoming of the symbolism of the *Cossists* was that it had no symbols for the coefficients of the equation. Instead they used such phrases as 'the number of things', meaning by this nothing other than the coefficient of the first power of the unknown. The symbolism was unable to express the coefficients in a general way.

In 1591 François Viète (1540–1603) published his *Introduction to the Analytic Art*. In this book Viète introduced the symbolical distinction between unknowns and parameters. He was the first to represent the coefficients of equations with letters. Viète used capital vowels A, E, I, O,

U, to represent the *unknowns* and the capital consonants B, C, D, F, G, ... to represent the *coefficients*. In addition, each quantity had a dimension: 1-*longitudo*, 2-*planum*, 3-*solidum*, 4-*plano-planum*, ... The dimension of each quantity was expressed by a word written after the symbol, thus for instance A *planum* was the second power of the unknown A while A *solidum* was the third power *of the same unknown*. Thus the letter indicates the identity of the quantity while the word indicates its particular power. This expedient makes it possible to use more than one quantity, and among other things makes it possible to express a substitution. Viète understood quantities as dimensional objects and therefore retained the principle of homogeneity. This principle stipulated that we can add or subtract only quantities of the same dimension, and that the result of this addition or subtraction will be a quantity of the same dimension as the operands. This is a carryover from geometry, because in geometry we cannot, for example, add length to volume. Even though Viète's symbolism is rather complicated, it was a great step forward. It was *the first universal symbolic language for the manipulation of formulas*. Viète was fully conscious of the importance of his discovery. He believed that this new universal method, or rather analytic art, as he called it, would make it possible to solve all problems. He ended his book with the words: '*Finally, the analytical art, having at last been put into the threefold form of zetetic, poristic, and exegetic, appropriates to itself by right the proud problem of problems, which is: TO LEAVE NO PROBLEM UNSOLVED.*' Thus he believed that his analytical art could solve all problems.

3. The Co-ordinative Form of the Language of Algebra: The Solution of an Equation as a Splitting of a Form (from Stifel to Euler)[5]

One of Cardano's merits was the systematic nature of his work. Therefore besides the equation of the form '*cubus and thing equal number*', the solution of which was discussed above, he presented rules for the solution of the other two forms of cubic equations. The rules for the solution of these equations are very similar in form to the first case, which can be obtained by simple substitutions. Therefore I will not discuss them here.[6] Instead, I would like to turn to a discovery Cardano made when he tried to

[5] The co-ordinative form and the compositive form are two forms of language, which we have found only in algebra. For more details see [Kvasz, 2005].

[6] The solution of the equation $x^3 = bx + c$ is $\sqrt[3]{\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 - \left(\frac{b}{3}\right)^3}} + \sqrt[3]{-\frac{c}{2} + \sqrt{\left(\frac{c}{2}\right)^2 - \left(\frac{b}{3}\right)^3}}$. The difference from the previous case seems to be small. The result is not a difference, but a sum of two cube roots, and beneath the quadratic roots (in Cardano's *binomium* and *apotome*) we have a difference instead of the sum of the two expressions. But these small differences have important consequences, because a difference of two positive numbers can become negative.

apply his general rule for the equations of the form '*cubus equals thing and number*' to the equation $x^3 = 7x + 6$. When he applied the rule he obtained a result we would express as:

$$x = \sqrt[3]{3 + \sqrt{-\frac{100}{27}}} + \sqrt[3]{3 - \sqrt{-\frac{100}{27}}}.$$

Below the sign for the square root a negative number appeared. The formula required him to find $\sqrt{-\frac{100}{27}}$, something he was not able to do. For the further progress of algebra it was crucial to understand what was going on when a negative number appeared below the square root sign. The rules of manipulation with formulas brought Cardano into a situation that was beyond his comprehension, a situation where he had to do something impossible. From the point of view of the projective form of language the square root of a negative number is a meaningless expression.

The way out of this situation led to a change of our attitude towards formulas. In the projective form, even if the reference of the language becomes indirect, the linguistic expressions are still understood as tools for representing some real objects. Nevertheless, the discovery of the *casus irreducibilis*, of the insoluble case, led to a gradual loosening of the bond between language and reality. The algebraic expressions are viewed more and more as *forms*,[7] as formal objects constructed from symbols, independent of any realistic context in which they are supposed to be interpreted. An important motive for a development in this direction was the situation in the theory of equations. Cardano considered equations such as $x^3 + bx = c$ and $x^3 = bx + c$ to be different problems, because he allowed only positive numbers for coefficients and solutions. For equations of the third degree this represents only a small complication, but in the case of equations of fourth degree we have seven different kinds of equations, and in the case of quintic equations fifteen. Now these kinds are not fully independent, because, as Cardano had shown, simple substitutions can transform an equation of one kind into another. Therefore it is rather natural to try somehow to reduce this complexity. It was Michael Stifel, whom we have already mentioned in connection with the introduction of the symbol for the square root, who first saw how this might be accomplished. In his book *Arithmetica integra* (1544) Stifel introduced rules for the arithmetic of negative numbers, which he interpreted as

[7] I use the word '*form*' in two different contexts. On the one hand I use it in the context of epistemology in phrases describing the different *forms of language*, thus for instance the projective *form* of language, the co-ordinative *form* of language, *etc*. On the other hand I use the word '*form*' in algebraic contexts, thus in phrases as *polynomial form*, *quadratic form*, *etc*. I hope that this will cause no difficulties for the reader.

numbers smaller than zero. That is an extension of the number concept, natural for the projective form of language, because negative numbers begin to play an important role as values of the auxiliary variables. But Stifel went further, beyond of the realm of the projective form of language, and started to use negative numbers also as coefficients of equations. This enabled him to unite all fifteen kinds of quintic equations, which formerly had to be treated separately, into one general *form*: $x^5 + ax^4 + bx^3 + cx^2 + dx + e = 0$.

Before Stifel attempted to solve an equation, he transferred all the expressions to one side, and in this way obtained the equation in a *polynomial form* $p(x) = 0$. Thus a polynomial as a mathematical object is first found in Stifel's work. He used a simple symbolism without symbols for coefficients. But the basic idea of reducing all the different cases to a single form, by allowing the coefficients to be negative, was decisive. A polynomial is an expression that co-ordinates different *formulas* into one universal *form*. Accordingly, we call the new form of language of algebra based on this approach the *co-ordinative* form. The universal form common to all cubic equations had remained hidden until then, because mathematicians bound the algebraic language too firmly to reality. Only when Stifel stopped observing a distinction between positive and negative coefficients did this new deeper unity become visible.

When we free ourselves from the understanding of the algebraic expressions as pictures of reality, and start to understand them as more or less independent formal objects, it becomes possible to accept the square roots of negative numbers simply as a special kind of expression. Even though we do not know what such expressions represent, we know how to calculate with them. This understanding is implicit in the book *Algebra*, written by Rafaelo Bombelli in 1572. Bombelli simply introduced rules for the addition, subtraction, multiplication and division of these new expressions, and did not ask what they stand for. Maybe the most pregnant expression of this view can be found in Leonard Euler's book *Vollständige Anleitung zur Algebra* from 1770, where imaginary numbers are called *numeri impossibiles*, because they are not smaller than zero, not equal to zero, and not greater than zero. Euler writes:

> They force themselves on our mind, they exist in our imagination and we have sufficient notions of them, because we know that $\sqrt{-4}$ means a number, which when multiplied by itself gives -4.

Thus even though in reality there is no quantity whose square is negative, we have a clear understanding of the meaning of the symbol $\sqrt{-4}$.

The transition from formulas to forms is also important for another reason. If we consider formulas as the basic objects of algebra, one central aspect remains hidden. A polynomial has many roots; thus in general there is not just one number satisfying the conditions of an algebraic problem. In the framework of the *perspectivist form of language* this was ignored, because in reality, in the normal case, the number of things we are looking for is unequivocally determined. Mathematicians therefore simply ignored the existence of other roots of an equation, and as the solution of the problem they accepted the root that made sense in the problem's context. In most cases they were not even aware that they were overlooking some solutions, because in most cases the other solutions were negative and thus—from the point of view of the perspectivist form of language—unacceptable. In connection with the *projective form of language* the situation was somewhat better. For the auxiliary equations it was necessary to take the negative solutions into account as well, because it can happen that a negative solution of the auxiliary equation corresponds to a 'true' (*i.e.*, positive) solution of the original problem. Nevertheless, as a solution of the whole problem mathematicians still accepted only a positive number, one that gave the 'number of things'. Only when the bonds tying the language to reality became looser did they accept that equations generally have more roots. Thus the transition from algebraic formulas to algebraic forms was crucial for the understanding of the relation between the degree of an equation and the number of its roots.

We expect a *formula* to tell us the result, to give an answer to the problem. A formula expresses a number we want to know, it represents the answer to the question we are asking. A *form*, on the other hand, is a function, giving different results for different inputs. It might not be easy to imagine that a given problem has more than one answer, because if we are asking something about reality, we expect that the answer is uniquely determined. Yet when we understand the equation describing the problem as a polynomial form, it becomes understandable that the form can produce the same value (usually zero) for more than one value of its argument. Thus the transition from formulas to forms makes it easier to accept that an equation can have more than one solution. Already Viète knew about this, because he had discovered the relations between the roots and the coefficients of an equation. But oddly enough all the roots he presents in his examples are positive. Only when the equation is understood as a form can the relation between the roots and the coefficients be disclosed in its entire generality, as was done independently by Albert Girard (1595–1632) and René Descartes.

The notion of a form attained its full meaning in the work of Descartes. From Euclid to Descartes a product of two quantities was understood as a quantity of a new kind. Thus a product of two segments was a rectangle, and by multiplying two lengths we obtained an area. The product of a rectangle with a segment gave a prism, thus a volume. It is true that the *Cossists* went beyond the three dimensions allowed in geometry, but the terms they used for the powers of the unknown show that they still thought of algebraic operations in geometrical terms. The influence of geometry is even more visible in Viète. Descartes left this tradition behind when he introduced a radically new interpretation of algebraic operations. For him for the first time the product of the segments x and y is not a rectangle, having the area equal to xy, but a segment having the length xy. This change might appear trivial, but it was one of the most important ideas in the development of algebra. When Descartes interpreted the product of two segments as a segment, he created a system of quantities closed under algebraic operations. Thus with a slight touch of anachronism we can say that *Descartes created the first example of a field*. By a field we understand a system of quantities containing 0 and 1 that is closed under the four algebraic operations and satisfies the commutative, associative, and distributive laws. From the epistemological point of view closure under operations means the grasping of the universe of discourse as a whole.

Thus algebraic language started to serve a fundamentally new function, the function of grasping the unity of the world represented by the language, grasping the way its different aspects are *co-ordinated*. This co-ordination exists on two levels. On the one hand we have the co-ordination of different formulas (for instance the different kinds of equations as *cubus and thing equals number* and *cubus equals thing and number*) into a single form of a polynomial. On the other hand we have a co-ordination of different quantities (for instance Viète's *longitudo*, *planum*, *solidum*, *plano-planum*, *etc*.) into a single field. Thus the language of algebra becomes a means for grasping the unity behind the particular formulas and quantities. This unity opens up a new view of equations. Instead of searching for a formula that would give us the value of the unknown, we face the task of finding all the numbers that satisfy the given form. In other words, we are searching for numbers which we can use to split the form into a product of linear factors. Consider, for example, the form

$$x^3 - 8x^2 + x + 42 = (x - 7)(x-3)(x + 2),$$

which shows us that 7, 3 and –2 are the roots of the polynomial $x^3 - 8x^2 + x + 42$. When we have found all roots, we are able to split the form $x^3 - 8x^2 + x + 42$ into linear factors $(x - 7)$, $(x - 3)$, and $(x + 2)$. This factorization shows that no other root can exist (for any number different from 7, 3 and –2 each factor gives a nonzero value and so their product is also nonzero). Thus the splitting of the form into linear factors gives a complete answer to the problem of solving a given equation. From the point of view of the co-ordinative form, to solve an equation means to *split a form into its linear factors*.

4. The Compositive Form of the Language of Algebra: The Solution of an Equation as a Resolvent (from Hudde to Lagrange)

After Albert Girard and René Descartes discovered that a polynomial of nth degree has precisely n roots, Cardano's 'formulas' had to be revised. These formulas determined only one root of the cubic equation. But a cubic equation has three roots, and so it became necessary to find a way to determine the remaining two roots. The Dutch mathematician Johann Hudde (1628–1704) found a procedure that makes it possible to find all roots of a cubic equation. Consider the equation

$$x^3 + px - q = 0.$$

Using the substitution $x = y - \frac{p}{3y}$ Hudde transformed this equation into

$$y^6 - qy^3 - \left(\frac{p}{3}\right)^3 = 0,$$

which was later named *Hudde's resolvent*. Despite the fact that it is a sixth-degree equation, it is simpler than the original equation, because by the substitution $y^3 = V$, we obtain a quadratic

$$V^2 - qV - \left(\frac{p}{3}\right)^3 = 0,$$

which has the roots

$$V_1 = \frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \quad \text{and} \quad V_2 = \frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

In order to get from the variable V back to y, we could take the cube root of V. Nevertheless, that would be a mistake, because in this way we would obtain only two roots of the original equation, one as the cube root of V_1 and the other as the cube root of V_2. But y is the root of an equation of the sixth degree, which has six roots. Hudde realized that the operation of root extraction is the step where we are losing roots. If in the equation $y^3 = V$ we substitute a number for the variable V, for instance 1, the equation $y^3 = 1$, as an equation of the third degree, has three roots. Thus besides the root $y = 1$, there are two others. These are the complex cube roots of unity. Their values are

$$\omega = -\frac{1}{2} + i\frac{\sqrt{3}}{2} \quad \text{and} \quad \omega^2 = -\frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

Here we used i for $\sqrt{-1}$. With the help of ω and ω^2 we can find all roots of Hudde's resolvent:

$$\begin{aligned} y_1 &= \sqrt[3]{V_1}, \quad y_2 = \omega\sqrt[3]{V_1}, \quad y_3 = \omega^2\sqrt[3]{V_1}, \\ y_4 &= \sqrt[3]{V_2}, \quad y_5 = \omega\sqrt[3]{V_2}, \quad y_6 = \omega^2\sqrt[3]{V_2}. \end{aligned}$$

The roots of the third-order equation, from which we started, will then be combinations of these roots:

$$x_1 = y_1 + y_4 = \sqrt[3]{V_1} + \sqrt[3]{V_2},$$
$$x_2 = y_3 + y_5 = \omega^2 \sqrt[3]{V_1} + \omega\sqrt[3]{V_2},$$
$$x_3 = y_2 + y_6 = \omega\sqrt[3]{V_1} + \omega^2 \sqrt[3]{V_2}.$$

We obtained each of the three solutions of the cubic equation in the form of a combination of two cubic roots, that is, in the same form as in Cardano's formulas. Nevertheless, the whole procedure is much more systematic, and the quadratic equation used to obtain the roots V_1 and V_2 is incorporated into the process of solution in a more symmetric way. Leonard Euler (1707–1783) tried to generalize Hudde's approach to the case of a general polynomial

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n = 0.$$

He wanted to find an analogous auxiliary equation, the so-called *Euler's resolvent*, the roots y_i of which would be connected to the roots of the original polynomial in a way similar to that in which the roots of Hudde's resolvent are connected to the roots of the original cubic equation. But Euler did not get very far, because the degrees of the resolvents he obtained were too high. For the cubic equation the degree of the resolvent was 6, for the biquadratic the degree of the resolvent was 24, and for the quintic, which was the main point of interest, the degree was 120. Euler could find no trick which would reduce the degree of the resolvent.

At that point Joseph Louis Lagrange (1736–1813) entered the discussion with a generalization of Euler's approach. Lagrange introduced a new kind of resolvent, which today bears his name. The *Lagrange resolvent* for a polynomial of the fourth degree does not have the degree 24, as Euler's resolvent did, but rather only the degree 3. Thus Lagrange's resolvent seemed to be a big step forward, because it made it possible to reduce the problem of solving an equation of the fourth degree to the problem of solving its resolvent, which was of the third degree. Lagrange succeeded in showing that all tricks for solving equations of the third and the fourth degree, presented in Cardano's *Ars Magna*, were also based on the use of resolvents. It is sufficient to look at the solution of the cubic equation, where an auxiliary quadratic equation (3) appeared. The difference is that, while in Cardano's solution we found this auxiliary equation by chance, in Lagrange's solution the resolvent appears in a systematic manner, and its importance is fully understood. Thus what was previously only a bit of good luck becomes with Lagrange a consciously applied method. All the techniques for solving an algebraic equation that had been discovered before Lagrange consisted explicitly or implicitly in the reduction of the initial equation to its resolvent, which was of lower degree. After gaining this insight into the process of solution of algebraic equations, Lagrange believed that the technique of reduction of an equation to its resolvent should also work for quintic equations, where the resolvent should be of the fourth degree. But here Lagrange met with a disappointment: the resolvent of the equation of the fifth degree turned out to be of the sixth degree. Six is much less than 120, which was the degree of Euler's resolvent, but it is still more than five, and so this resolvent is of no help in solving the original equation. Lagrange was thus confronted with a strange situation. His method worked nicely for the equations of the third and fourth degree, where the resolvents had lower degrees than the initial equations, but in order to solve the equation of the fifth degree, his method required the solution of an equation of the sixth degree.

But even if the methods of Hudde, Euler, and Lagrange do not work for quintic equations, they do have something important in common. All three of them are trying in some way or other to *compose* the solution x of the original problem from the solutions y of the resolvent and the complex roots ω of unity. Of course with the advantage of hindsight we see here the first step towards the notion of factorization. But leaving this aside for the moment, we can say that the *compositive form of language* is based on the idea of decomposing a problem into independent parts that are easier to handle. Thus instead of alternative perspectives on the same problem, which was the basis of the projective form of the language of algebra, in the compositive form we have to do with a variety of analogous problems arranged in such a way that the solution of the more complicated problem can be reduced to the solution of the preceding simpler problems.

5. The Interpretative Form of the Language of Algebra: The Solution of an Equation as the Construction of its Splitting Field (from Lagrange to Gauss)

The interpretative form of language is connected with the full acceptance of the square roots of negative numbers as standard mathematical quantities. The acceptance of these so-called complex numbers amounts to a reification of another layer of the language of algebra. Nevertheless, this reification was very different from the previous one connected with the birth of algebraic symbolism. In the case of algebraic symbolism, specific acts were reified. For instance, the operation of taking the cube root of a number was turned into a symbol representing the result of this operation, or the operation of addition was turned into a symbolic representation of the sum (*i.e.*, of the result of this addition). By contrast, in the case of the interpretative form, the reification does not consist in turning the results of some additional operations into objects. It is impossible to take the square root of a negative number. Thus, strictly speaking, there is nothing to be reified. We have at our disposal no concrete activity, no performance, that we could declare to be a new object. A square root of a negative number instead represents the impossibility of an operation, a failure of the language. The process that was brought to its completion by the construction of the model of the complex plane consisted therefore in something fundamentally different from simply turning operations into objects.

This process started with overcoming the antipathy towards linguistic expressions containing square roots of negative numbers and simply adding them to the language of algebra. The square roots of negative numbers were considered to be a special kind of expression, for while their use was well understood, it was not clear what they signified. Already in the perspectivist form of language of algebra new expressions, the negative numbers, appeared that also had no direct reference. Thus it seemed reasonable to consider the complex numbers as something similar to the negative numbers, that is, expressions of the language whose reference is indirect. Nevertheless, such an interpretation of the complex numbers is doomed in advance to failure. A negative number can be turned into a positive one by means of an appropriate substitution, and so may be regarded as a representation of a representation. In the case of a complex number nothing like this is possible. A complex number can never indicate any number of things, directly or indirectly, because the complex numbers cannot be linearly ordered, or in Euler's words '*they are neither greater nor smaller than nothing*'.

Another proposed attempt at an interpretation of the square roots of negative numbers was based on the idea of ascribing to them not an indirect but rather a subjective meaning. Descartes introduced the term *imaginary number* for the square root of a negative number and from this term the contemporary notation i for $\sqrt{-1}$ is derived. In his *Geometry* he writes that in the case of an equation that does not have enough true (*vraies*, *i.e.,* positive) and false (*fausses*, *i.e.*, negative) roots, we can imagine (*imaginer*) some further, imaginary (*imaginaires*) roots. Nothing in reality corresponds to the imaginary roots. Descartes introduces them only in order to ensure that an equation of the nth degree will have exactly n roots. Euler in his *Algebra* has this to say about the square roots of negative numbers: '*they are neither greater nor smaller than nothing; but neither are they nothing, and that is why we must consider them as impossible. In spite of that they present themselves to our reason (Verstand), and they find place in our imagination (Einbildung); that is why they are called imaginary (eingebildete) numbers. But even though these numbers, as for instance* $\sqrt{-4}$, *are according to their nature absolutely impossible, we have a sufficient notion of them, because we know that by this a number is indicated, which multiplied by itself gives* *the result* -4*; and this notion is sufficient in order to subsume these numbers under computations.*' [Euler, 1770, p. 61]. Thus for Euler too these quantities exist only in our imagination. But this subjective interpretation of the complex numbers cannot explain how it is possible for computations involving these non-existent quantities to lead to valid results about the real world. It is as if a biologist, after reflecting on centaurs, were able systematically to bring forth new knowledge about horses, and have his claims substantiated by biology. If the complex numbers make it possible to disclose new knowledge about the world, they must be related to the real world in some way. A purely subjective interpretation is therefore rather unsatisfactory.

In 1799 Carl Friedrich Gauss (1777–1855) created a geometric model of the complex numbers. The problems with the interpretation of the complex numbers were not solved by finding a way to ascribe to an individual complex number some reference in the context of the particular problem, in the solution of which the expression of that complex number appeared. Instead, all possible expressions to which we cannot ascribe a reference were reified. Rather than seeking interpretations of individual complex numbers in particular contexts, a model is constructed for all of them at once. Gauss's complex plane was thus perhaps the *first model in the history of mathematics*. For the first time an artificial universe of objects was constructed in which the language as a whole is interpreted. Thus the idea of a model is to be met with in Gauss's work some 70 years before Beltrami. From the epistemological point of view, Gauss's model of complex numbers is rather similar to Beltrami's model of non-Euclidean geometry. First of all, in both cases the model serves the purpose of making a doubtful theory acceptable. Before Gauss the complex numbers had a dubious status, just as before Beltrami the status of non-Euclidean geometry was not clear. The fundamental similarity between Gauss's plane and Beltrami's model lies in that both of them *actualize a whole world*. Thus when Beltrami constructed a model of the non-Euclidean plane inside a circle, he actualized the whole world of non-Euclidean geometry. In a similar way, Gauss's plane represents the whole world of complex numbers. At first sight it might seem that some regions must escape when he represents the complex numbers by means of a plane, because the plane he uses cannot be viewed as a whole. But we must remember that Gauss's model concerns not geometrical lines but algebraic expressions. Therefore even if from the geometrical point of view Gauss's plane escapes our field of vision, from the algebraic point of view it nevertheless represents the whole world—because the world of algebra is a world of operations. Therefore to represent the whole world of algebra does not mean to be '*entirely in the field of vision*' (as it is in geometry), but to be '*closed under operations*'. And Gauss showed that the sum, difference, product, and quotient of two points of the complex plane is again a point of this plane. Therefore Gauss's plane is a closed universe, in which we can interpret all algebraic operations. A third similarity between Gauss's and Beltrami's models consists in the fact that in both of them a *translation is incorporated into the language*. Gauss ascribes to each complex number a point of the plane, and he shows how is it possible to translate the algebraic operations of addition or multiplication of complex numbers (*i.e.*, operations of the internal language of the model) into the language of geometric manipulations with the points of the plane (the external language of the model). On the other hand Beltrami ascribes to every figure of the non-Euclidean plane a figure inside a circle of the Euclidean plane, and he shows how is it possible to translate the notions of non-Euclidean geometry (*i.e.*, expressions of the internal language of the model) into the language of Euclidean geometry (the external language of the model).

With the help of his model Gauss proved the *fundamental theorem of algebra*, which says that every polynomial of the nth degree has n roots. In proving this theorem Gauss showed that the problem with the equations of the fifth degree, for instance the equation $x^5 - 6x + 3 = 0$, does not lie in the fact that they lack roots. He showed that in the complex plane there are exactly five points corresponding to solutions of a quintic equation—*i.e.*, points whose co-ordinates, when substituted into the equation, return the value zero. Thus the problem with the equations of the fifth degree turns out to be much more subtle. Even though the roots of such equations exist, *it is impossible to express them by the means of the language of algebra*. That is, there is no general formula formed from integers, the four arithmetic operations ($+$, $-$, $\times$, $\div$) and root extraction ($\sqrt[5]{\ }$, $\sqrt[17]{\ }$, $\sqrt[542]{\ }$, ...), which would represent the roots of the above equation. To see more clearly what the problem is, let us imagine a huge sheet of paper, on which all the algebraic formulas are already written. Thus on our paper we have all formulas consisting of 500, 1 000, 1 000 000 or any other number of symbols. We would like to prove that none of the five roots of the equation $x^5 - 6x + 3 = 0$ is to be found on this piece of paper. How can we prove such a claim? It is not difficult to prove that when we add, subtract, multiply or divide any two algebraic formulas (excepting division by 0), we obtain again an algebraic formula. This means that all the formulas contained on our huge sheet of paper form a closed system, which is in algebra called a *field*. We want to show that this field does not contain any root of the above-mentioned polynomial. Here we see the immediate advantage the interpretative form of language yields. By reifying the whole world of algebraic expressions a *modal predicate*, that something is impossible to do (*i.e.,* an equation cannot be solved) is turned into an *extensional predicate*, that some numbers do not belong to a field. Due to this reification of the whole world of algebraic expressions we begin to see what the problem might be with equations of the fifth degree. The problem is not that these equations lack solutions, but rather that the language of algebra is not rich enough to express these solutions. Thus in a sense the reification of the two worlds, Gauss's reification of the world of complex numbers and the above-described reification of the world of algebraic formulas, makes it possible to understand the phenomenon of insolubility. The insolubility of the equations of the fifth degree means that their roots (which as complex numbers do exist) fall outside the world of algebraic formulas. Thus the language of algebraic formulas and the world of roots of algebraic equations do not fit together at all points. But even if the interpretative form of language were able to understand what the problem with quintic equations might be, it does not have the means to prove that the general quintic is insoluble. To be sure, some simpler problems—simpler, that is, from the point of view of the fields involved—like the impossibility of trisecting an angle, could be handled at this stage. But in order to prove that quintic equations are indeed insoluble, it was necessary to reify the next layer of the language of algebra and in this way to create a much stronger tool—group theory.

6. The Integrative Form of the Language of Algebra: The Solution of an Equation as the Factorization of the Galois Group (from Gauss to Galois)

In the framework of the interpretative form of language of algebra the whole world of algebraic formulas was reified. The world of algebraic formulas is a field, that is, a system of objects closed under the four basic arithmetical operations. Nevertheless, closer investigation reveals that inside this world there is a whole range of different subfields, a whole range of smaller worlds, from the smallest one, the field Q of all rational numbers, to some slightly bigger fields like $Q(\sqrt{2})$ to the greatest field of all, the complex numbers C. Gauss showed that in the field of all complex numbers each polynomial of the nth degree has n roots, and so the polynomial $x^5 - 6x + 3$ has five roots. The field of numbers that can be expressed by algebraic formulas lies somewhere between the fields Q and C. It is richer than the field Q of all rational numbers because it contains irrational numbers such as $\sqrt{2}$. On the other hand it is poorer than the field C of all complex numbers, because the number π is not expressible by any algebraic formula. In order to show that the roots of the equation $x^5 - 6x + 3 = 0$ cannot be expressed by any algebraic formula, we have to characterize more precisely which numbers can be expressed by such formulas. But the interpretative form of language is not up to this task. That form is able to reify a whole world and turn it into an object (a field), and it is able to describe the transition from one such world to another as a translation. But it cannot compare different worlds—because it is always restricted to one abstract structure. Thus, for instance, in the construction of the Gaussian model of the complex numbers we are able to translate algebraic operations with complex numbers into geometrical transformations of points of the Gaussian plane, but we are able to do this only because the structures of the complex numbers and the Gaussian plane are isomorphic. In a way, therefore, we are always working with the same abstract structure, and move only between its different realizations. The formal relations are the same in both cases and so we can say that the interpretative form is able to reify only one structure, and move this structure from one medium into another.

Only through the transition to the integrative form of language, which replaces the translation between equivalent structures (isomorphism) by an embedding into a richer structure (homomorphism) does it become possible to compare different structures. And again, as in the previous form of language, there is a whole range of analogies between how Cayley and Klein introduced the integrative form of language in geometry (see [Kvasz, 1998, p. 154]) and how Evariste Galois (1811–1832) introduced the integrative form of language in algebra. The first common feature is the existence of a *neutral basis*, a fundamental level of description in terms of which all the structures are to be compared. In geometry the projective plane was such a basis, and the different geometrical structures were compared as structures introduced into this neutral basis. In algebra the field of complex numbers is such a neutral basis, and all the fields Galois was working with are subfields of this field. The next common feature is the role played in both cases by *group theory*. In geometry, Klein compared the different geometries by comparing the transformation groups associated with them. In algebra, group theory is the basic means which makes it possible to compare the different fields. Thus it can be said that the integrative form of language integrates different (geometric or algebraic) worlds by embedding the symmetries of these worlds (transformation groups or groups of automorphisms) into one neutral structure (the projective plane or the field of all complex numbers). Nevertheless, the proof of the insolubility of equations of the fifth degree is a bit too complex from the technical point of view and I have therefore decided to split it up into its basic steps in order to make it more comprehensible.

6.1. *The Epistemological Interpretation of the Notion of a Group*

Let us first consider a general equation of the third degree

$$x^3 + ax^2 + bx + c = 0. \qquad (5)$$

From Gauss we already know that this equation has three roots $\alpha_1, \alpha_2, \alpha_3$. These three roots exist as three points in the complex plane. Using the three numbers α_1, α_2, α_3 we can create a world associated with equation (5), namely the smallest field which contains all three roots. This field is usually represented by the symbol $Q(\alpha_1, \alpha_2, \alpha_3)$. It is the field that results when we add the three numbers $\alpha_1, \alpha_2, \alpha_3$ to the rational numbers along with everything else needed to ensure that the new system is closed under the four operations (thus for instance $5\alpha_1 + 7\alpha_2$ and similar combinations). So far we are not interested in whether this field can be constructed by algebraic means. Of course we already know the answer, because we know Cardano's formulas, and so we know that the three roots can be expressed by formulas. But for the moment we will ignore this fact, and we will study the field without reference to this question—because we would like to apply the knowledge we gain to the cases where no explicit formulas are known.

The world of algebra is a constructed world that emerged through the systematic reification of algebraic operations. Therefore we can consider the epistemic subject of algebra to be the subject who performs these operations. This subject is not identical to any particular mathematician, because a mathematician can make mistakes while the epistemic subject is connected rather with the way that operations *should* be executed. In geometry the epistemic subject had the form of a viewpoint, and in algebra I will by analogy speak about a 'viewpoint of the blind'.[8] A 'blind human' is also situated in his world. Nevertheless, he is usually not fixed in his world in an unambiguous way. This resembles geometry, where the viewpoint was also not given in a fixed sense, but could be moved together with a simultaneous shift of the horizon. After the world of all algebraic formulas was reified in the form of a field, the epistemic subject was not situated in this field in an unambiguous way. Its position is fixed only formally, by means of important objects like 0 and 1. In the field Q of all rational numbers the *zero* indicates where the subject '*stands*', while the *one* indicates the positive orientation and the length of his '*steps*' (if we are allowed to use these corporeal analogies). These two objects are clearly distinct, because $0 + 0 = 0$, while $1 + 1 \neq 1$, therefore the 'blind

[8] As mentioned earlier, I maintain that the language of algebra emerged from the reification of performative acts. The epistemic subject of this language is thus the subject who performs these acts. When I call him blind, I do not mean real blindness. I only want to indicate that in order to understand this epistemic subject it is useful to leave out of consideration our usual way of orientating ourselves in the world by means of vision. The world of algebra is not open to sight. It is rather the world of motoric schemes. When a blind man learns to move around in a building, for instance in a school, he memorizes all the possible movements he can make, and the building is represented in his mind as a structure constituted by these movements. Thus for instance the classroom might be represented as the result of 'go 12 steps straight, there you will find a staircase, go two stories up, turn to the left, walk along the wall and the first door on the left is your classroom'. The blind have to spend some time learning all the necessary movements in the building. Afterwards, they can move about freely and safely.

man' (or the epistemic subject of the language of algebra) can distinguish them. As soon as he learns to identify these two objects, he can employ algebraic operations to construct (and so also identify) every rational number. In the world of rational numbers his position is unambiguously determined. Unfortunately, this field is too small, and it can help us to solve only very simple equations.

This situation changes radically when we enrich the world of the 'blind man' by adding the three numbers α_1, α_2, α_3, and turn from the field Q to the field $Q(\alpha_1, \alpha_2, \alpha_3)$. In this new field a fundamental problem appears: These and only these numbers satisfy the equation (5) and because of this the 'blind man' can distinguish them from all others. Yet the 'blind man' cannot discriminate the three numbers α_1, α_2, α_3. He knows that there are three of them, and he knows that they are different, but he cannot tell which is which. To their numerical values, which determine the position of these three numbers on the complex plane, the 'blind man' has no access. Algebra allows only a finite number of steps in a calculation, while the determination of the numerical value of the roots of algebraic equations requires in general an infinite number of steps, an approximation procedure. From the algebraic point of view the three numbers α_1, α_2, α_3 are, at least initially, indiscernible. The notion of a group was introduced in order to express this indiscernibility formally. The fact that the quantities α_1, α_2, α_3 are indiscernible means that we can change their order without affecting the field $Q(\alpha_1, \alpha_2, \alpha_3)$. This field is the reification of a fragment of the language of algebra, a fragment which is invariant under any permutation of the three numbers α_1, α_2, α_3. We say, that the field $Q(\alpha_1, \alpha_2, \alpha_3)$ is symmetric with respect to such permutations.

We can visualize this symmetry by imagining a displacement of the 'blind man' in his world $Q(\alpha_1, \alpha_2, \alpha_3)$ that he cannot detect. The 'blind man' cannot distinguish the three numbers. It may happen that he thinks he has them in front of himself in the order α_1, α_2, α_3, when in reality they lie before him as α_2, α_3, α_1. It is not difficult to see that he can err precisely in $3{\cdot}2{\cdot}1 = 6$ ways, namely (writing just the subscripts):

$$(1,2,3), \quad (1,3,2), \quad (2,3,1), \quad (2,1,3), \quad (3,1,2), \quad (3,2,1). \tag{6}$$

Thus the field $Q(\alpha_1, \alpha_2, \alpha_3)$, corresponding to the equation (5) of the third degree, has six symmetries. These symmetries can be combined. For instance we can, after exchanging the first two roots (creating the order (2, 1, 3)), exchange the first and the third (resulting in the order (3, 1, 2)). It is interesting to notice that the result will be different if we make these changes in the opposite order. If we first exchange the first and third roots, yielding (3, 2, 1), and then exchange the first and second roots, the result will be (2, 3, 1).

The symmetries of a given field form a closed system under the operation of composition, which is called a *group*. A group is something like a field; it is a system of objects closed under specific operations. The only difference is that while a field is formed by numbers and is closed under arithmetic operations, the objects forming a group are not quantities, but reified transformations which are closed under composition. From the epistemological point of view, a *group is a reification of a further layer of algebraic operations*, the layer of symmetries of a field. When dealing with a group we have to do with operations on two levels. On the one hand, operations are the very objects that form a group, and on the other we have the operation of their composition. Thus we could say that a group is a closed system of operations with operations.

6.2. *The Symmetry Groups of Fields Belonging to Solvable Equations*

After a short detour towards the notion of the group, let us return to the question of solvability of equations. We already know that the field $Q(\alpha_1, \alpha_2, \alpha_3)$, which corresponds to the equation of the third degree, has at most six distinct symmetries. In order to see them more clearly, it will be useful to introduce the distinction between permutations and substitutions, which goes back to Augustin-Louis Cauchy (1789–1857). A *permutation*, which we will write as for instance (1, 3, 2), represents a symmetry of the field $Q(\alpha_1, \alpha_2, \alpha_3)$ in a reified form. Thus different permutations represent the different orders in which the roots $\alpha_1, \alpha_2, \alpha_3$ may be arranged. On the other hand substitutions, which we will write as $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, represent the same symmetries, but now in a nonreified way, as operations. This expression means that the first root stays where it is, while the other two exchange places. Thus in general the symbols in the upper line indicate the roots that are moving, while the corresponding symbols in the lower line indicate their destinations. The symbols (1, 3, 2) and $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ represent the same symmetry, the former in reified form as the result of a transformation, the latter as the transformation itself. Clearly, to each permutation there corresponds precisely one substitution, and conversely. The difference between these two notions is only an epistemological one. Nevertheless, for the birth of group theory this difference played an important role.

From the fact that one substitution corresponds to each permutation, we know that in the field $Q(\alpha_1, \alpha_2, \alpha_3)$ there are six substitutions. Thus we have a reified and a nonreified version of the group of symmetries of the field. And this made possible a very clever trick. Galois asked, what would happen if we were to *apply a particular substitution to all permutations*. If, for instance, we apply the substitution $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ to the permutation (2, 1, 3), the substitution indicates that 2 will be turned into 3, 1 into 2, and 3 into 1. The result will be the permutation (3, 2, 1). Galois reified the permutations, combined them to form a system, and then investigated what would happen with this system if he applied the same substitution to all permutations. If we take the permutations

$$(\mathit{1},\mathit{2},\mathit{3}),\quad (1,3,2),\quad (\mathit{2},\mathit{3},\mathit{1}),\quad (2,1,3),\quad (\mathit{3},\mathit{1},\mathit{2}),\quad (3,2,1),$$

and apply to them the same substitution $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, the result will be

$$(\mathit{2},\mathit{3},\mathit{1}),\quad (2,1,3),\quad (\mathit{3},\mathit{1},\mathit{2}),\quad (3,2,1),\quad (\mathit{1},\mathit{2},\mathit{3}),\quad (1,3,2).$$

At first sight it may seem that the permutations have simply changed places. Nevertheless, the surprising fact is that the three permutations printed in italic characters changed their positions only among themselves, and the other three permutations again only among themselves. Thus it seems, that the permutations can be divided into two blocks:

$$(\mathit{1},\mathit{2},\mathit{3}),\ (\mathit{2},\mathit{3},\mathit{1}),\ (\mathit{3},\mathit{1},\mathit{2})\quad \text{and}\quad (1,3,2),\ (2,1,3),\ (3,2,1).$$

Galois discovered that substitutions can accomplish only one of two things: either they rearrange the permutations in the blocks, while leaving the blocks intact (like the substitution $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$), or they can exchange whole blocks. But no substitution can mix the elements between the blocks. So, for instance, no substitution can shift the permutation (1, 2, 3) into the second block, while leaving the remaining two permutations (2, 3, 1) and (3, 1, 2) in the first block. No permutation can break the borders of the blocks. Either the blocks stay put or they move as wholes. A substitution can never move only a part of one block into another. And Galois discovered that this respecting of the boundaries of the blocks of permutations by the substitutions is closely related to the fact that the equation of the third degree is solvable. The reason for this is that the three roots α_1, α_2, α_3 can be expressed by the use of the four arithmetic operations plus root extraction. Any field created in this way has symmetries which can be divided into blocks, and these blocks into further blocks, and so on, so that at the end we will come to a block, the number of elements of which is a prime number (in our case we got a block with three elements after the first step). In other words the symmetry group of fields, which corresponds to solvable equations, can be factorized into cyclic factors.

6.3. *The Insolubility of the Equation of the Fifth Degree*

By reifying the symmetries of the particular fields Galois reached a level of abstraction that allowed him to understand why equations of the fifth degree are in general insoluble. Gauss had already shown that every equation of the *n*th degree has *n* roots. Therefore each quintic equation has five roots—α_1, α_2, α_3, α_4, α_5—, and the problem is only that these roots cannot be expressed by algebraic means. The integrative form of language makes it possible to understand why this is so. It enables us to replace the question whether the roots α_1, α_2, α_3, α_4, α_5 can be expressed by algebraic means by the question whether the group of symmetries of the field $Q(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ can be split into smaller and smaller blocks. This last question is not so difficult. It is sufficient to take all the permutations of five elements (of which there are $5! = 120$), and see what happens to them under different substitutions. The case of five elements is more complicated than the case of three elements, which we discussed above, but these difficulties are not fundamental. Galois discovered that the only possible division into blocks is a division into two blocks containing 60 elements each. But if we restrict ourselves to one of these two blocks, we have a group with 60 elements, which is one of the most interesting groups in mathematics. It is called the *alternating group of five elements*. This group cannot be further divided into blocks, because the permutations mix the elements between any blocks. The discovery of this fact was one of the most surprising moments in the history of algebra.

The symmetry group of every field that is constructible by algebraic means can be factored into a system of nested blocks. Galois discovered that the symmetry group associated with such an 'innocent' equation as $x^5 - 6x + 3 = 0$ cannot be factored in such a way. That means that no field constructed by algebraic means can ever contain the roots of this equation. Thus there cannot be any general formula for the solution of fifth-degree equations analogous to Cardano's formulas for cubic equations. This shows that the solvability of equations in terms of radicals is a rather exceptional phenomenon. Only equations with special symmetry groups turn out to be solvable. Cubic equations, for instance, are solvable because the associated fields only have six symmetries, given by the permutations listed above. These permutations can be divided into two blocks which in a sense correspond to the symmetries of Cardano's formula. Beginning with the equation of the fifth degree, however, no such division is possible, and therefore there is no formula capable of solving this equation. The universe of algebraic formulas is too simple. It does not allow us to construct fields with symmetries complex enough to encompass the field $Q(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ associated with the fifth-degree equation. Therefore algebra has to shift its focus from formulas to *algebraic structures*. Algebraic structures, as for instance groups, decide what can be formally expressed and what cannot. The discovery of the alternating group of five elements marks the start of modern structural algebra.

7. Concluding Remarks

Let us end the story here. I took up the problem of the solution of algebraic equations as a kind of thread to lead us through the labyrinth of the history of algebra. Instead of going further I would like to summarize our results. We have discriminated six forms of language of algebra, which differ in the way they conceive of a solution of algebraic equations. To solve an equation means:

1. To find a *regula*, *i.e.*, a rule written in ordinary language, which makes it possible to *calculate* the 'thing', that is, the root of the equation.
2. To find a *formula*, *i.e.*, an expression of the symbolic language, which makes it possible to *express* the root of the equation in terms of its coefficients, the four arithmetical operations, and root extraction.
3. To find a *factorization* of the polynomial form, *i.e.*, to *represent* the polynomial form as a product of linear factors.
4. To find a *resolvent*, *i.e.*, to *reduce* the given problem, by means of a suitable substitution, to an auxiliary problem of a lesser degree.
5. To find a *splitting field*, *i.e.*, to *construct* the field that contains all the roots of the equation.
6. To find a *factorization* of the Galois group of the splitting field, *i.e.*, to *decompose* the group of automorphisms of the field into blocks.

Besides these differences on the *intentional* level, the particular forms of language differ also with respect to their *ontology* and *semantics*. Thus their discrimination can be seen as a first step towards a better appreciation and understanding of the richness of the philosophical issues that we encounter in algebra. The process of gradual reification of operations described in the paper can be viewed as a contribution to the discussion about the ontology of mathematics. It seems that each form of the language of algebra has an ontology of a specific kind. To develop a philosophical account of mathematical ontology would therefore require an account of the common aspects as well as of the differences among the ontologies of the particular forms of language.

References

AL-KHWĀRIZMĪ, M. [850]: Matematíčeskije traktaty. Russian translation by A. P. Jushkevich and B. A. Rosenfeld. Taškent: FAN, 1983.

CARDANO, G. [1545]: *Ars Magna, or the Rules of Algebra*. Cambridge, Mass.: MIT Press, 1968.

DESCARTES, R. [1637]: *La Géométrie*. Appendix to *Discours de la Méthod*, Leiden. *The Geometry of René Descartes*. D. E. Smith, trans. New York: Dover, 1954.

EUCLID: *The Thirteen Books of the Elements*. Thomas Heath, trans. New York: Dover, 1956.

EULER, L. [1770]: *Vollständige Anleitung zur Algebra*. Leipzig: Reclam, 1911.

KLEIN, J. [1934]: *Greek Mathematical Thought and the Origin of Algebra*. Cambridge, Mass.: MIT Press, 1968.

KVASZ, L. [1998]: 'History of geometry and the development of the form of its language', *Synthese* **116**, 141–186.

—— [2000]: 'Changes of language in the development of mathematics', *Philosophia Mathematica* (3) **8**, 47–83.

—— [2004]: 'The invisible dialog between mathematics and theology', *Perspectives on Science and Christian Faith* **56**, 111–116.

—— [2005]: 'Similarities and differences between the development of geometry and of algebra', in C. Cellucci and D. Gillies, eds, *Mathematical Reasoning and Heuristics*. London: King's College Publications, pp. 25–47.

SCHOLZ, E., ed. [1990]: *Geschichte der Algebra*. Mannheim: Wissenschaftsverlag.

STEWART, I. [1973]: *Galois Theory*. London: Chapman and Hall, 1992.

STRUIK, D. J., ed. [1969]: *A Source Book in Mathematics*. Cambridge, Mass.: Harvard University Press.

TOTTI RIGATELLI, L. [1996]: *Evariste Galois*. Basel: Birkhäuser.

VIÈTE, F. [1591]: *Introduction to the Analytical Art*, in [Klein, 1934], pp. 313–353.

VUILLEMIN, J. [1962]: *La Philosophie de l'Algèbre*. Paris: UFP.

VAN DER WAERDEN, B. L. [1980]: *A History of Algebra, from al-Khwārizmī to Emmy Noether*. Berlin: Springer.