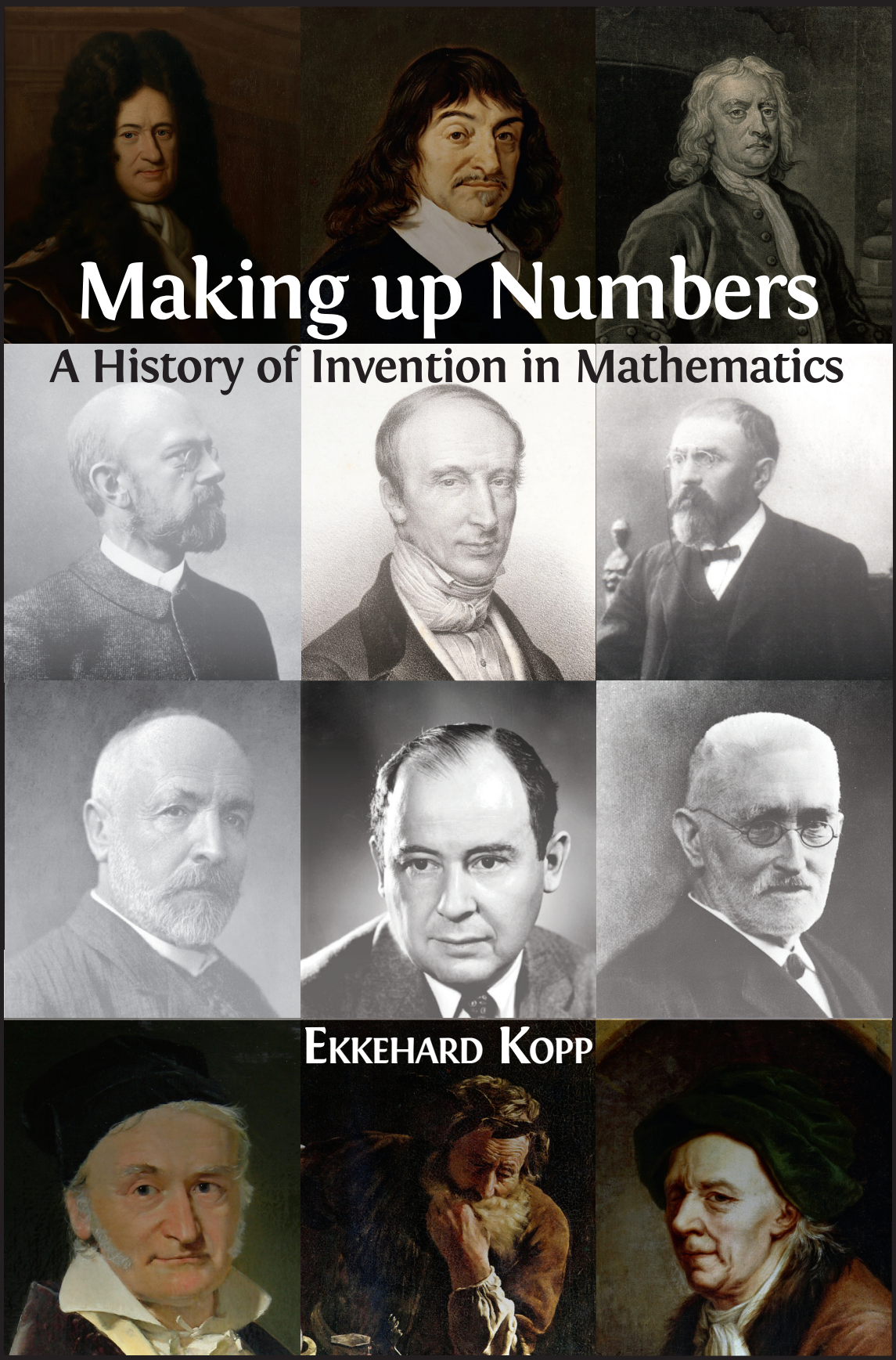

Making up Numbers

A History of Invention in Mathematics

Ekkehard Kopp

https://www.openbookpublishers.com

ISBN Paperback: 978-1-80064-095-5

ISBN Hardback: 978-1-80064-096-2

ISBN Digital (PDF): 978-1-80064-097-9

DOI: 10.11647/OBP.0236

Cover images from Wikimedia Commons. For image details see captions in the book.

Cover design by Anna Gatti.

CHAPTER 2

Writing and Solving Equations

..to invent is to discover that we know not, and not to recover or resummon that which we already know.

Sir Francis Bacon, *The Advancement of Learning*, 1605

Summary

We review the development and acceptance of our current decimal system of number symbols, known to historians as *Hindu-Arabic* numerals. This reflects its Indian origins as well as its further development in the Arab caliphates that conquered the Middle East, parts of Central Asia, North Africa and Spain in the eighth and ninth centuries. In addition, while also taking initial steps towards *algebra* in a systematic study of quadratic and cubic equations, Arab scholars were crucial in the preservation and translation of Greek manuscripts.

The Hindu-Arabic numerals (together with most classical Greek mathematics) remained largely unknown in Europe until the twelfth century. A key figure in its transmission was Leonardo of Pisa (or *Fibonacci*), whose influence led to the replacement of reliance on the abacus by the use of Arabic numerals, with intermediate steps recorded on paper. This practice spread quickly, first in commerce, where symbolic notation served as shorthand. This, together with the study of equations, led to 'formulae' for the solution of cubics and quartics, first published in 1545 in the influential *Ars Magna* by Girolamo Cardano.

1. The Hindu-Arabic number system

While classical Greek texts were highly influential in cementing the dominance of geometry as the principal domain of certainty and proof, quite different sources influenced the development of arithmetical techniques and the symbolic representation of numbers. One key example was the early development, primarily in India, of the *decimal number system* we all take for granted today. However, piecing together this history today is complicated by an almost total lack of primary written sources, so that much of what is known is based on secondary sources.

 https://doi.org/10.11647/OBP.0236.02

A brief summary of results from recent scholarship on this topicis given in [**25**].[1] The classic text [**31**] describes in greater detail the gradual evolution, from origins in the Brahmin period in India (third century BCE) to early modern Europe, of the number symbols we use today. The earliest appearance of the nine number symbols that preceded the number digits $1, 2, ..., 9$ of our decimal system is found in decrees inscribed on pillars during the reign of King Ashoka (third century BCE). Over the next millennium the symbols were gradually transformed – together with the addition of zero by a dot (denoting 'absence') – although number symbols for $10, 20, ..., 90$ also remained in use, while combinations of symbols were initially needed to depict higher numbers. The earliest known Indian mathematician, *$\overline{A}$ryabhata* (b.476) had a system of names for powers of ten, for example. The individual symbols for the nine digits were further developed by Arab mathematicians from the eighth century onwards, introduced to Spain during its Muslim occupation and transmitted via Italy to the rest of mediaeval Europe.

The crucial advantages of a place-value system using only nine symbols were realised quite early on. From at least 2000 years ago, columns on Chinese counting boards represented different powers of 10. It has been speculated that trading between the Chinese and Hindu cultures in South-East Asia may have led to an exchange of ideas, culminating (probably around 600) in the 'Hindu' numerals for numbers beyond 9 being dropped in favour of a full place-value system, including using the nine digits, together with a dot (later a circle) for zero. Evidence for this suggestion includes an inscription found in Cambodia dated to the year 683, shown as the 605th year of the Saka period and displaying the symbols then used for 6 and 5, separated by a dot. Similarly, an early eighth-century Chinese astronomical work explicitly describes the 'Hindu' use of a place-value system, including use of the dot, and comments that it made calculation 'easy'.

The earliest extant fragment referring to this 'Hindu system' comes from a Syrian priest, *Severus Sebokht*, who commented as early as the year 662 that the Hindus had a valuable calculation method 'done with nine signs'. He did not refer to the dot.

In any event, the most significant impact of Indian mathematics on modern mathematical techniques really began with the transmission of these numerals to the Arab world, apparently dating from about 770, when an Indian scholar visiting Baghdad showed his hosts a Sanskrit text containing calculation methods based on a decimal place system using nine number symbols and a symbol indicating zero (or 'absence'). Their use in the development of arithmetical techniques, including square root extraction, clearly impressed Islamic scientists, although it took a considerable time for the new methods

[1]This nomenclature 'Hindu-Arabic number system' has traditionally been used to describe aspects of mathematical development and innovation originating in the Middle East and/or the Indian subcontinent. Current debates on more appropriate terminology continue.

to gain wide acceptance in practice. The Sanskrit text was translated into Arabic, providing the basis of what gradually became widely known and eventually prized as *'Hisāb al-Hind'* ('Indian calculation').

As argued in [**25**], it seems likely that, while Indian mathematicians worked with a well-developed decimal place-value system for integers by the early eighth century, first steps in the extension of the system to include decimal fractions were due to somewhat later Arab mathematicians (perhaps beginning with al-Uqlidisi in the tenth century). Over the next few centuries Arab mathematicians also gradually modified the numerals, bequeathing to Europe what became the familiar number symbols of today's decimal system.

More generally, in the eighth and ninth centuries, a wide variety of scientific and mathematical texts was gathered together in Baghdad during the reign (786-809) of the caliph *Harun al-Rashid* and translated into Arabic. The different techniques and notations arising from these disparate sources were gradually harmonised and developed further. His influential *House of Wisdom*, became a public academy under his son, *al-Ma'mun* (reigned 813-33), who brought together Islamic as well as other scholars, often from Jewish, Zoroastrian or Christian backgrounds.

Mathematical research flourished in an atmosphere of free intellectual enquiry not dissimilar to that of the Greek city states. Although these researches were undertaken under a monotheistic regime, for at least two centuries secular enquiry was not regarded as creating conflict with religious dogma, and was openly encouraged. Nonetheless, such openness to new ideas was never fully universal throughout the caliphate and sadly it did not last. By the eleventh century the atmosphere had changed drastically. Greater religious orthodoxy in the *madrasas* focused more and more exclusively on Islamic law, to the effective exclusion of 'foreign sciences'. From that time onwards, mathematical activity other than basic 'practical' arithmetic, was increasingly discouraged, although some important exceptions remained, as described below.

Fortunately, by this time Europe was beginning to emerge from a prolonged period of turmoil and became receptive to the knowledge being transmitted, at first piecemeal but increasingly effectively, by scholars visiting the Arab world. European scholars discovered the work of Arab mathematicans such as *al-Khwarizmi* (ca. 780-850 AD) and his successors, who had adopted and gradually modified the numerals from India into a superior symbolic number system. The English word *algorithm* stems from a Latin transcription (*Algoritmi*) of his name. His most influential text deals with the simplification of equations, and our word *algebra* is a corruption of *al-jabr*, which describes the practice of transposing a term from one side of an equation to the other (with the attendant change of sign).

The earliest extant Arabic text (by *al-Uqlidisi*) shows that, by 962 AD, this number system included the symbol 0 and had a clear place value notation, as well as taking (still somewhat incomplete) steps towards handling decimal fractions. The author extols the flexibility and convenience of the 'Hindu' notation, and points to its advances over earlier 'finger-bending' methods to denote numerals, as well as to the ease of recording and checking complex calculations in the new system.

The final steps towards a fuller understanding of the decimal system, and of the convenience afforded by using the decimal notation when calculating with fractions, developed relatively slowly. However, treatises discussing approximations to fractions whose denominators include numbers other than 2 or 5 (so that the decimal expansion does not terminate), appeared in the Arab world in the twelfth century, showing that the authors clearly understood that the approximation can be made arbitrarily close by continuing far enough, providing *'an infinite number of answers, each being more precise and closer to the truth than the preceding one'*. The full decimal notation, with a vertical line separating the integral and fractional parts, first appears around 1500, in the work of *al-Kashi.* By then, Europe was catching up fast, with similarly sophisticated use of both finite and infinite decimal expansions, starting with the work of *Francois Viète* and *Simon Stevin* in the 1570s. All these developments played a fundamental role in creating the decimal number system we have today.

1.1. Arabic algebra. The great merit of the contributions of the Islamic mathematicians was not only that they preserved and transmitted the classical Greek works. They combined systematic geometrical methods developed by Euclid and his successors with earlier *ad hoc* methods inherited from the ancient Babylonians on linear and quadratic equations. In this, they constructed algebraic methods by which equations could be solved, yielding (as al-Khwarizmi already observes) *numbers* as the result of their calculations. Geometric arguments then served to *justify* the methods used—something which appears to be absent from Babylonian 'algebra'.

As noted above, the term 'algebra' stems from the term *al-jabr,* popularised in the famous work of al-Khwarizmi, *'Al-Kitāb al-muhtasar fī hisāb al-jabr wa'l muqabāla* ('The Compendious Book on the Calculation of *al-jabr* and *muqabāla*'), written around 825.[2] In his writings al-Khwarizmi does not make explicit use of equations, but describes the terms and steps in his solutions verbally – he was apparently unaware of the work of Diophantus.

Key to his solution of problems leading to quadratic equations of the form

$$ax^2 + bx + c = 0$$

[2]More details can be found in **[6]**.

was the classification of problems involving an unknown, its square and constants (what he would call 'squares, roots and constants') into six basic categories as follows (the modern symbolic notation is given in brackets):

squares equal roots	$(ax^2 = bx)$
squares equal numbers	$(ax^2 = c)$
roots equal numbers	$(bx = c)$
squares and roots equal numbers	$(ax^2 + bx = c)$
squares and numbers equal roots	$(ax^2 + c = bx)$
roots and numbers equal squares	$(bx + c = ax^2)$

Note that this enables him to avoid negative coefficients, as well avoiding 0 on the right. His justification for the solution techniques he uses is always given in geometric constructions – essentially the 'completion of the square' that we discussed when looking at Babylonian cuneiform problems. For these, he compares two geometric figures, so that all quantities concerned must remain strictly positive.

A solution recipe (or 'algorithm') was then given for each of the six types. For the fourth (squares plus roots equal numbers) he gives the following example (here taken from [**6**]), where $a = 1, b = 10, c = 39$, giving the equation $x^2 + 10x = 39$:

You halve the number of roots, which, in this problem, yields five, you multiply it by itself, result is twenty-five; you add it to thirty-nine; the result is sixty-four; you take the [square] *root, that is eight, from which you subtract half of the root, which is five, The remainder is three, that is the root of the square you want, and the square is nine.*

In our notation his general solution of $x^2 + bx = c$ is therefore

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

which is exactly what we saw in our earlier Babylonian example. Unlike the Babylonians, however, al-Khwarizmi takes great care to produce a geometric proof – in fact, he produces two proofs, one of which is virtually identical to our square-completion in Figure 5.

The influence of the 'six problems of al-Khwarizmi' is clear from subsequent texts throughout the Islamic period and early European mathematics – his solution methods were learnt by rote, and apparently more complex problems were systematically reduced (both by al_Khwarizmi and his successors) to one of the six types. He compared various combinations of 'squares, roots and numbers' and modified them by means of *al-jabr* (typically by moving terms to be subtracted to the 'other side', where they would be added), or *muqabāla* ('compensating'), which was done by reducing an apparently more complex equation by grouping like terms on the side where

the net result is positive. For example, $5 + x^2 = 3x + 12$ is simplified to $x^2 = 3x + 7$. He also simplified both sides by cancelling common factors, finally arriving at one of the six types. Throughout, his comparisons were of aggregates of quantities that could be represented geometrically by means of squares and rectangles, so that the *dimensions* of the figures represented on each side remained the same.

From these beginnings, various generations of Islamic mathematicians fashioned systematic methods which they could justify geometrically in the Euclidean manner. After the translation of Diophantus' *Arithmetica*, late in the ninth century, their work went beyond the quadratic, to solve certain types of problems that we would see as leading to cubic equations and beyond. Although geometry remained the principal means of justifying their increasingly complex techniques and results, by the time of *Omar Khayyam* (1048-1131), Islamic mathematicians had studied and classified more than a dozen types of problems that we would today describe by means of cubic equations. In addition, in their studies of the salient features and differing advantages of the Babylonian sexagesimal number system (used throughout astronomy and astrology) and the new 'Hindu' decimal system, they made advances in unifying techniques for the manipulation of whole numbers and fractions, leading, in practice, to greater freedom in handling them both as 'numbers', even if they never fully articulated a consistent conceptual framework for these techniques.

Their ideas and techniques were to be taken up and ardently pursued by Renaissance mathematicians in Europe.

2. Reception in mediaeval Europe

In Europe, the fall of the Roman Empire in the fifth century resulted in its replacement by local feudal systems, led by often barely literate barons, who carved out local fiefdoms and engaged enthusiastically in local military campaigns. Latin survived primarily in Italy and what is now Southern France, where the Roman notion of the *quadrivium*—a term apparently coined by the Roman scholar *Boethius* (ca. 480-524) to describe the study of arithmetic, geometry, music and astronomy—was still regarded as necessary for the educated man, if only in a residual form with much-attenuated content.

During the reign of *Charlemagne* (742-814) in Central Europe the newly established multi-ethnic *Holy Roman Empire* began to develop a new focus on learning, based largely in the monasteries, under the leadership of *Alcuin of York* (735-804). He combined the quadrivium and the *trivium* (grammar, rhetoric and logic) into a comprehensive curriculum. The mathematical manuscripts available to Alcuin were few: the principal mathematical

text circulating widely in Europe at this time was Boethius' *De Institutione Arithmetica,* a less than perfect version of an introductory text by the first century neo-Pythagorean *Nichomachus.*

The classical Greek works were to remain unknown in Western Europe until the twelfth century. Moreover, by the end of the ninth century, the brief revival of learning under Charlemagne had itself become overwhelmed amidst internal strife and by various invasions, from the East by Magyars, from the North by Vikings and from the South by Saracens. However, many of the monastic schools established by Alcuin survived these onslaughts and a more permanent revival began at the turn of the millennium, with *Gerbert d'Aurillac* (ca. 945-1003), who became Pope in 999. He introduced the Hindu-Arabic numerals on a counting board whose columns represented positive powers of 10, with zero marked by an empty column.

His grasp of this system may have been imperfect, but his efforts heralded the introduction, less than a century later, of new techniques, recently rediscovered from many manuscripts distributed by Arab scribes throughout the regions of the Islamic conquest. A motley group of translators was especially active in the Spanish city of Toledo, which had been retaken by Christian forces from its Moorish rulers in 1085. This military victory provided scholars with access to a multitude of Arabic manuscripts, including translations of Greek scientific and mathematical classics. Many of the earliest Latin versions of these manuscripts were produced via translation from Arabic into Hebrew by scribes from the city's substantial Jewish community, who were fluent in Arabic as well as Latin.

From these local beginnings, translations of Arabic manuscripts obtained from a variety of sources were soon to produce a substantial body of mathematical literature, available in Latin, and widely transmitted to scholars throughout Europe.

2.1. Fibonacci. A key figure in this early period of transmission was *Leonardo of Pisa* (ca.1170 to ca.1250), now more commonly known as *Fibonacci* ('son of Bonaccio'), although this nickname is probably a nineteenth century invention. His *Liber Abaci,* published in 1202, was highly influential. In his youth, Leonardo was taught mathematics in Bugia on the Barbary Coast (now in Eastern Algeria), which was then part of the Western Muslim Empire. He travelled widely throughout the Muslim world, becoming familiar with Euclid's *Elements* and the Greek methods of geometric proof, as well as with the Hindu-Arabic numerals, the decimal place system, and the algebraic approach to solving equations of al-Khwarizmi. Upon his return to Pisa he joined the academic court of the Holy Roman Emperor, Frederick

II, writing several influential texts – of which the practically oriented*Liber Abaci* was by far the most successful and widely read.[3]

Despite its title, the first part of the book focuses on the way in which the Hindu-Arabic numerals provide an *alternative* to mechanical calculation. Since Roman times, practical calculations had usually been performed in Europe with an *abacus* (typically, a wooden frame strung with wires carrying different coloured beads as counters) or similar mechanical device; the final answer was then written down in Roman numerals. Leonardo showed how these mechanical devices could be by-passed by recording on sheets of paper, in Hindu-Arabic notation, the various steps and results of applying simple *algorithms*for combining numerals when adding, subtracting, multiplying or dividing.[4] These algorithms remain essentially unchanged in modern primary school curricula. In other words, *Liber Abaci* became the text for performing calculations *without* the abacus.

Leonardo begins with the numerals:

The nine Indian figures are 9, 8, 7, 6, 5, 4, 3, 2, 1. With these nine figures, and with the sign 0, which the Arabs call ***zephir*** *(cipher), any number whatsoever is written, as is demonstrated below.*

The bulk of the book is devoted to a wide range of practical problems in mensuration, commerce and currency conversion, as well as developing algebraic techniques to handle a wide range of linear and quadratic equations. In these, and in describing methods for series summation, he gives meticulous geometric justifications of his methods, in the style of Euclid. Notably, in the later chapters he readily uses negative numbers as solutions to certain equations and calculates accurately with these. He justifies rules for adding and multiplying positive and negative numbers, although this is always done in the context of 'practical' calculation, as in the following example, found in Chapter 13 of the *Liber Abaci*.

The problem states that four men find a purse, and that for each, the sum of his original wealth plus the purse, is in a simple proportion in relation to the original wealth of the next two, in a circular pattern: the first plus the purse will have wealth *double the sum of the second and third*, the second plus the purse *triple that of the third and fourth*, the third plus the purse *quadruple that of the fourth and first*, the fourth plus the purse *quintuple that of the first and second*. He shows that this problem can only

[3] Astonishingly, this highly influential Latin text was not translated into any modern language for 800 years – an English edition [**42**] finally appeared in 2002!

[4] The process of producing paper from woodpulp dried into thin flexible sheets was invented in China early in the second century. It was gradually transmitted via the Middle East (especially Baghdad), reaching Western Europe by the thirteenth century, and was soon produced in local paper mills, displacing the earlier uses of papyrus and vellum. The original name for paper in Europe, *bagdadikos*, indicates its tranmission to Europe via the Arabic world.

be solved if *the first man has a debit* (that is, he owes money!) which means that the solution requires negative numbers. The problem is indeterminate in general, but he finds the smallest solution: the four men have original wealth $-1, 4, 1, 4$ respectively, and the purse is 11. These numbers provide a solution, because

$$\begin{aligned}(-1+11) &= 10 = 2(4+1)\\ (4+11) &= 15 = 3(1+4)\\ (1+11) &= 12 = 4(4+(-1))\\ (4+11) &= 15 = 5(-1+4).\end{aligned}$$

Leonardo does not write -1 as a number, but expresses it as a 'debt' (what we might call 'negative equity'!).

2.2. The Fibonacci sequence. The final chapter of *Liber Abaci*, and his subsequent *Liber quadratorum* (1225) demonstrate clearly that Leonardo is comfortable with the full range of Islamic algebra, including the solution the general quadratic and certain cubic equations. His handling of now well-known number theory results like the Chinese Remainder theorem (finding numbers that leave pre-assigned remainders when divided by a fixed set of primes) foreshadows *number theory* that would be developed by *Fermat* some 350 years later. Yet he is most widely known for a seemingly much more trivial result, which has fixed his name in modern public consciousness: the *Fibonacci sequence.* This appears innocently in Chapter 12 of the *Liber Abaci:*

A certain man had one pair of rabbits together in a certain enclosed place, and one wishes to know how many are created from the pair in one year when it is in the nature of them in a single month to bear another pair, and in the second month those born bear also.

The (unspoken) assumption here is that the first pair are new-borns at the start of the year in question and that rabbits begin to mate when one month old. This assumption ensures that we treat all the pairs equally. At the end of the first month the first pair is therefore still the only one, but by the end of the second month they have borne their first pair of offspring. The sequence therefore begins with

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ..$$

since, from the third term onward, each term is the sum of the preceding two: by the end of the second month there are two pairs, in the third month the adults produce another pair (so now there are 3 pairs), in the fourth month both they and their firstborn produce pairs (making the total 5 for the fifth month) and so on. By the end of the twelfth month (assuming all the rabbits survive till then) we have 144 pairs.

While this is an old and rather trivial problem—the sequence appears in various Indian mathematical writings as early as the sixth century—it gains

mathematical interest when one considers the successive ratios:

$$\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \frac{21}{13}, \frac{34}{21}, \cdots$$

These ratios can be shown to come ever closer to the *golden ratio* $\frac{1}{2}(1+\sqrt{5})$. The term "golden ratio" seems to have been coined as late as the nineteenth century. However, the number itself generated much excitement in artistic circles, especially during the Romantic era.[5]

To see how this limiting ratio is found, we denote the nth Fibonacci number by f_n. Hence the Fibonacci sequence starts with $f_1 = f_2 = 1$, and $f_{n+1} = f_n + f_{n-1}$ for $n = 2, 3, \ldots$. The ratios $(r_n)_{n\geq 1}$ of successive terms satisfy the identity

$$r_n = \frac{f_{n+1}}{f_n} = \frac{f_n + f_{n-1}}{f_n} = 1 + \frac{f_{n-1}}{f_n} = 1 + \frac{1}{r_{n-1}}.$$

If we accept (for a proof see *MM*) that the $(r_n)_n$ do indeed *converge* ('get ever closer to') to some number x as n grows, then the limiting value x of the two sequences $(r_n)_{n\geq 1}$ and $(r_{n-1})_{n>1}$ must clearly be the same. Therefore the relationship $r_n = 1 + \frac{1}{r_{n-1}}$ will, for large n, approximate the equation $x = 1 + \frac{1}{x}$, which we can write as $x^2 - x - 1 = 0$. The usual formula for the solution of a quadratic equation provides the positive value

$$x = \frac{1+\sqrt{5}}{2} = 1.6180339987\ldots$$

Since we know from the Theodorus lesson that $\sqrt{5}$ is not a rational number, it follows that neither is x, the golden ratio. This again raises the question how one should *define* these quantities consistently.

In geometric terms, the 'golden ratio' (also called the 'golden section') has a history that predates Leonardo by at least 1500 years. Greek geometers (possibly even in Pythagorean times) were concerned to determine the point C on a straight line segment AB such that 'the whole is to the greater part as the greater is to the smaller'. Euclid called this the division of AB in *mean and extreme ratio*. By this he meant that the lengths AC, BC should satisfy the proportion

$$AB : AC = AC : BC,$$

(which makes AC the *mean proportional* between AB and BC, as discussed in Pythagorean music theory) and finding C requires the square root of their product, as

$$AC^2 = AB.BC.$$

[5]More recently, much has been published about the apparent ubiquity of the Fibonacci sequence and the golden ratio in nature, be it in the shapes of snail shells, configurations of flower petals, seed heads, pine cones, etc. A guide to the underlying mathematics of these phenomena can be found in the classic text *Introduction to Geometry* by HMS Coxeter. A less technically challenging reference is *Fascinating Fibonaccis* by Trudi Hammel Garland.

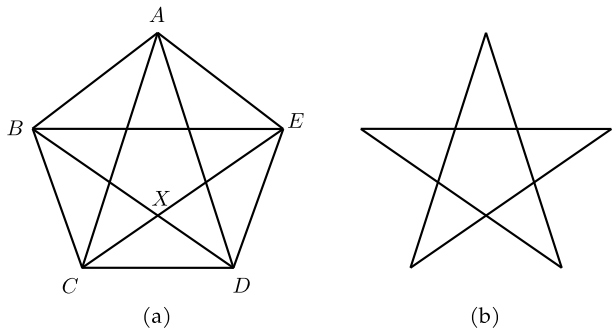

Figure 13. Pentagon and pentagram

Geometrically, AC is the side of a *square* whose area is the same as that of the *rectangle* with sides AB, BC.

To find C we put the 'greater part' $AC = a$, the 'lesser part' $BC = b$, making 'the whole' $AB = a + b$. The mean and extreme ratio is then

$$\frac{a+b}{a} = \frac{a}{b},$$

so it satisfies $\frac{a}{b} = 1 + \frac{b}{a}$. Setting $x = \frac{a}{b}$, this produces the identity $x = 1 + \frac{1}{x}$ satified by the limit of the Fibonacci ratios, so that $\frac{a}{b} = \frac{1+\sqrt{5}}{2}$. Remember, however, that $\frac{a}{b}$ is a ratio of *lengths* and that a, b cannot be taken as multiples of a chosen unit, since this identity has no solution in whole numbers.

The golden section is easy to construct geometrically. It arises as the ratio between the diagonal and side of a *regular pentagon*, i.e. a pentagon whose sides and internal angles are all equal. This fact suggests that its origins may well lie in Pythagorean times. The regular pentagon and the pentagram – the five-pointed star comprising the diagonals, which some claim may have been used by the Pythagoreans as a sign of recognition – see **[25]**) are pictured in Figure 13.

In Figure 13(a) the regular pentagon $ABCDE$ has diagonals AC, AD, BD, BE and CE, and if X is the point of intersection of BD and CE, the triangles BDC and CDX are similar, so that the corresponding sides are in proportion: for example, the ratios of lengths $\frac{BD}{CD}$ and $\frac{CD}{XD}$ are equal. Hence $CD^2 = BD.XD$. But clearly $CD = BC = BX$, so $BX^2 = BD.XD$ so that X divides BD in mean and extreme ratio. The pentagram is shown in Figure 13(b).

2.3. Maestri d'abbaco. Leonardo was clearly a highly accomplished mathematician, yet, apart from occasional glimpses of originality, his work is

chiefly a well-rounded compilation of earlier discoveries by the Greeks, Hindus and Arabs. In bringing these different strands together, however, he and the other translators of the twelfth and thirteenth century stimulated the revival of mathematics in Europe, especially in Italy, France and (a little later) in England and in the many Germanic statelets of Central Europe. While the tensions created by problems whose solutions included negative numbers and roots of equations that could not be expressed as ratios of whole numbers became increasingly apparent, it would take several generations before mathematicians were fully persuaded of the need to accept such solutions as *numbers*.

Through the success of Leonardo's *Liber Abaci*, the discipline (*abbaco* in Italian) of replacing the abacus and Roman numerals with the algorithms for calculating with Hindu-Arabic numerals, soon spread widely throughout much of Western Europe, especially through the growing influence of mercantile cities like Pisa, Genoa, Venice and Amalfi. In place of barter and direct trade, the international trading companies of these cities employed more sophisticated practices such as bills of exchange, promissary notes, letters of credit and loans, all of which required double-entry bookkeeping and arithmetical skills that were *not* part of the quadrivium taught in the early universities of Bologna, Paris or Oxford. The problem-solving techniques displayed in the *Liber Abaci* led to a new cadre of mathematics teachers, who became known as *maestri d'abbaco*, meeting the needs of the merchant class. Pen and paper calculation based on the decimal system began to replace the traditional counting boards, and a large collection of problem-solving manuals were produced to instruct budding merchants.

The innovations these *maestri* introduced were largely notational: the Arabs and Leonardo had described their calculations using words rather than symbols, but the pressure of teaching meant that, for the *maestri,* abbreviations (for square, cube, etc) soon became a common practice. The symbols for unknowns we find so familiar were still some way off, however. Typically, an unknown (which we might denote by x) was known as 'the thing' (*cosa),* and what we would write as $x + \sqrt{y}$, a fourteenth century instructor would typically describe as *a thing plus a root of some quantity.* An 'equation' might be constructed verbally, involving the unknown and its square and some numbers, and manipulated, according the rules already described by al-Khwarizmi, into one of his six standard forms. Gradually, problems being considered would include some aspects that seemed divorced from practical (commercial) use. The increase in complication in the expressions being explored led to the use of abbreviations, such as c for the 'cosa' or unknown, ce for its square, 'censo', and cu for its cube, 'cubo'. Higher powers, e.g. $ce\ cu$ ('censo di cubo') also began to appear. At this stage these abbreviations remained just that: no apparent meaning as *numbers* was attached to them.

The introduction of further symbolic notation was prominent in the work of the German *Coss* tradition, which had adapted the practices of the Italian *maestri* by the early sixteenth century. The influential treatise *Die Coss*, published in 1525 by *Christoff Rudolph* (1499-1545) contains a full list (in Germanic script) of symbols for powers up to the ninth and makes full use of signs such as $+$ and $-$ as well as a square root sign, to shorten the complicated expressions he discusses. (See **[6]** for a more detailed discussion.)

In practice, the consistent use of symbols helped to speed the acceptance of quantities that lacked geometric representation. However, Rudolph, who clearly recognised the link between what we would call the arithmetic progression $0, 1, 2, 3, 4, ...$ and its geometric counterpart $1, 2^1, 2^2, 2^3, 2^4 ...$, did not take the seemingly obvious next step of using the symbol 2^k for the kth member of the latter progression.

This advance had already been made in France by *Nicolas Chuquet*, whose precise dates are somewhat uncertain. He was possibly the first to use zero and negative numbers as exponents. His unpublished 1484 manuscript, *Triparty en la Science de Nombres* (rediscovered only in the late nineteenth century) was partly reproduced, without acknowledgement, in a 1520 textbook by *Etinenne de la Roche*. Chuquet invented an essentially modern notation to describe arbitrary integral powers and also (as we saw in the **Prologue**) introduced names for ever higher powers of 10. He used positive exponents for positive powers of the unknown, such as $.8.^3$ (which we would read as $8x^3$) and denoted negative exponents by adding m after the exponent, as in $.7.^1m$, which is our $7x^{-1}$. He recognised that multiplication of these two terms involves multiplication of the coefficients and *addition* of the exponents, resulting in $.56.^2$ (in our terms, $56x^2$). For addition and subtraction he employed the abbreviations also used by *Luca Pacioli* (1445-1517): $\overline{p}, \overline{m}$ for 'plus' and 'minus'. Our 'equals' sign $=$ first appeared in the 1557 *Whetstone of Witte* by *Robert Recorde* (1512-1558), who also introduced $+, -$ to Britain.

Throughout the sixteenth century, the boundaries between the arithmetical techniques inspired by the *maestri* and the study of the Greek classics at universities gradually became blurred. Newer editions of Euclid's *Elements* began to utilise arithmetical descriptions of Euclidean results, especially in Book X, where Euclid classifies various types of incommensurables (see **Chapter 1**, Section 4). His difficult geometric constructions become more transparent if the outcomes are presented as surds. Thus calculation using rationals and irrationals together no longer appeared as foreign. Calculations with surds such as $(2 - \sqrt{3}) \times (2 - \sqrt{3})$ also drew attention to the need for consistent rules when multiplying signs. Formal rules for sign

multiplication featured prominently in Pacioli's widely-read 1494 publication, *Summa de arithmetica geometria proportioni et proportionalita.* This contained a systematic account of the techniques pioneered by the *maestri* and *cossists*, pointing to their practical as well as theoretical utility.

Despite these advances, resistance to the acceptance of the reality of negative solutions to problems as *numbers* proved more difficult to overcome. I pick up this story in the next chapter, but will turn first to further progress in the solution of equations.

3. Solving equations: cubics and beyond

While progress with algebraic symbolism was steady rather than spectacular, in the first half of the sixteenth century algebraic techniques for solving polynomial equations beyond the quadratic took a major step forward. Publicly, the catalyst was the publication of the *Ars Magna* (1545) by *Girolamo Cardano* (1501-1576). The story behind this highly influential book is quite convoluted. The breakthrough was probably made in the early 1520s, but remained buried in secrecy for two decades. An unfortunate consequence was that the mathematician primarily responsible for advancing the subject at this time, *Scipione del Ferro*, died in 1526 and was largely forgotten by his peers and in many subsequent historical accounts. This has also meant that no direct evidence remains of the process by which he arrived at his results, although the final version of the techniques is well-documented.

For some time after the work of Leonardo of Pisa, the analysis of cubic equations had remained firmly within the Arabic tradition of addressing the problem geometrically, rather than algebraically. This tradition boasted a major work by *Omar Khayyam,* the polymath perhaps best known for a selection of about 1000 poems given the title *The Rubaiyat* by its Victorian translator, Edward FitzGerald.

Omar Khayyam's best-known mathematical treatise, *On the Proofs of the Problems of Algebra and Muqabala,* classifies fourteen different types of 'cubic equation' (he actually lists 25 cases, but 11 of these reduce to become linear or quadratic) with *positive* solutions. In each case he shows how to construct the solution of the cubic equation geometrically. His solutions invariably require the construction of a *conic section*. Specifically, this is either a circle, an ellipse, a parabola or a hyperbola, formed by the intersection of a plane with a circular cone (see Figure 14 in **Chapter 3**). The last three cannot be constructed by straightedge and compass alone—we will discuss their Ancient Greek origin in the next chapter.

Omar Khayyam insists that powers higher than cubes, such as the *square-square* of Diophantus, do not exist 'in reality' although he states that they constitute 'theoretical facts'. He argues that algebraic methods can be used

to solve cubic equations, but that proofs dealing with the third power require solid geometry.

Once again it is much more convenient to outline his methods in algebraic terms. Today we write the general cubic equation in the form

$$x^3 + ax^2 + bx + c = 0,$$

since we can always divide by the (non-zero) coefficient of x^3 if we need to.

However, Khayyam worked geometrically and insisted on writing the coefficient of the linear term in x as a square and the constant term in the equation as a cube: this would enable him to maintain the 'three-dimensional nature' of each of the terms on the left. The equation is therefore written as

$$x^3 + ax^2 + b^2x + c^3 = 0.$$

We express his method in modern notation: the substitution $x^2 = 2py$ (which defines a *parabola* centred at the origin) turns the equation into

$$2pxy + 2pay + b^2x + c^3 = 0,$$

which Khayyam recognises as defining a *hyperbola*. He has to deal separately with different cases for a, b, c, any of which can be positive, negative or zero, and he restricts himself to cases where the solution (the point of intersection of the parabola and hyperbola in question) has positive coordinates. While he works entirely in a geometric setting, he treats his solutions, whether rational or irrational, as providing *numbers,* rather than as geometric magnitudes.

3.1. Cardano's formula. This, more or less, remained the state of affairs for cubic equations as inherited by *Scipione del Ferro* around 1500. In Italy, as familiarity with decimal notation and algebraic manipulation grew, the problem of finding the solutions of various classes of cubics by means of an *algebraic* formula, rather than through a geometric construction, had begun to receive much attention. At this time also, the solution of particular equations – still expressed rhetorically – lent itself to a kind of public mathematical jousting which gradually became a popular sport in some cities. Scholars would challenge each other to solve problems in public, with bystanders engaging in bets on the outcome. Skilled combatants saw these challenges, and the associated betting, as a way of making a living from their craft.

In one such contest in 1535, *Niccolo Fontana* (1499-1557), an able scholar nicknamed *Tartaglia* (The Stammerer) roundly defeated the less talented *Antonio Fior*. It seems that the latter had learnt how to solve one class of cubic equations when studying as a pupil of del Ferro. As became apparent somewhat later, del Ferro had probably managed to find solution methods for all three classes into which cubics with positive solutions had been divided at

the time. Tartaglia, possibly aware of del Ferro's success, had managed, independently, to solve one class of cubic equations. In a short burst of intense creative activity just before the contest, he was also able to master the solution method for the class of equations Fior had learnt from del Ferro. He now set problems that Fior could not solve, while solving all thirty equations that Fior had set for him. The contest caused a minor sensation.

Hearing of this, Cardano approached Tartaglia, and, after much effort, persuaded the reluctant 'Stammerer' to divulge his techniques. Although Tartaglia, wishing to protect his livelihood, had sworn him to absolute secrecy, Cardano was already in the process of constructing his *Ars Magna* as a definitive algebra text and decided to break his promise. He tried to make amends by clearly citing del Ferro and Tartaglia as the authors of the solution methods for cubics featured in the book.

However, when the *Ars Magna* appeared, a furious Tartaglia accused Cardano of plagiarism and treachery. It was too late to repair the damage. A bitter dispute ensued, in which Cardano's principal defence was that the late del Ferro, rather than Tartaglia, had been the first to solve all three types of cubic equation. Cardano and his gifted student *Lodovico Ferrari* (1522-1565) claimed that they had confirmed this when they consulted del Ferro's papers, kept by yet another of del Ferro's students in Bologna, in 1543.[6]

Although Tartaglia published his book *Quesiti et Inventione Diverse* the following year, setting out his methods in verse, in coded form, Cardano's clear exposition and comprehensive range, presented in his *Ars Magna,* had already brought him fame that overshadowed Tartaglia's.[7] In the ensuing controversy, Cardano was ably supported by Ferrari, who had by then discovered a general solution method for quartic equations (see below), which Cardano also included in his *Ars Magna.* In a much-delayed challenge contest with Tartaglia in 1548, Ferrari energed a clear victor. An embittered Tartaglia died in obscurity and poverty nine years later.

One can appreciate the scale of the advances made by del Ferro and Tartaglia by reading Cardano's text. As usual, it is much easier for us to understand the method using modern algebraic notation. The wording of Cardano's verbal solution is reminiscent of solutions described on Babylonian clay tablets or by al-Khwarizmi. Here is an extract for comparison before we discuss his solution methods in modern terminology.

Cube the third part of the number of 'things' , to which you add the square of half the number of the equation, and take the root of the whole, that is square root, which you will use, in the one case adding the half of the number which you just multiplied by itself, in the other case subtracting the same half, and you will have a 'binomial'

[6]Extensive extracts from the correspondence between all these combatants are presented in [**12**], including the extract from Cardano's *Ars Magna* reproduced below.

[7]Tartaglia's book is available at http://www.it.wikisource.org/wiki/Quesiti_et_inventione_diverse. See also [**6**]

and 'apotome' respectively; then, subtract the cube root of the apotome from the cube root of the binomial, and the remainder of this is the value of the 'thing'.

Note the use of the Euclidean terms 'binomial' and 'apotome' (see **Chapter 1**, Section 3) to describe the sum and difference of the two terms under the cube root. Cardano provides an example, using the cubic $x^3 + 6x = 20$. He states this equation verbally as: *'the cube and* 6 *'things' equals* 20*'*, and leaves his solution in the form

$$\sqrt[3]{(\sqrt{108}+10)} - \sqrt[3]{(\sqrt{108}-10)}$$

which he does not attempt to simplify further, although $x = 2$ clearly also solves the equation.

Even in modern notation, the algebraic manipulations we need to arrive at Cardano's solution are somewhat more technical than those used so far – readers in a hurry may skip the two shaded sections below without much loss of continuity.

The general cubic equation

$$x^3 + ax^2 + bx + c = 0$$

can be reduced to an equation of the form $y^3 + py = q$ for appropriate constants p, q, so that the quadratic term is elimininated (following a pattern encountered already in the work of Diophantus). Simply set $y = x + \frac{a}{3}$, so that $x = y - \frac{a}{3}$. Expressed in terms of y, the equation takes this form, with $p = b - \frac{a^2}{3}$ and $q = -(c + \frac{2a^3}{27} - \frac{ab}{3})$.

To solve $y^3 + py = q$, first notice that for *any* A, B we have

$$(A-B)^3 = A^3 - 3A^2B + 3AB^2 - B^3,$$

so

$$(A-B)^3 + 3AB(A-B) = A^3 - B^3.$$

Thus, if we can find A, B to satisfy $3AB = p$ and $A^3 - B^3 = q$, then $y = A - B$ satisfies $y^3 + py = q$, and $x = y - \frac{a}{3}$ solves the original equation.

To find such A, B for the given values of p, q, note that $B = \frac{p}{3A}$ will lead to $A^3 - \frac{p^3}{(3A)^3} = q$, which reduces to a *quadratic* equation in A^3, namely

$$(A^3)^2 - qA^3 - (\frac{p}{3})^3 = 0.$$

The formula for the general quadratic now provides the solutions

$$A^3 = \frac{q}{2} \pm \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}$$

(although Cardano only takes the positive square root) and therefore

$$B^3 = -\frac{q}{2} \pm \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}.$$

Finally, Cardano writes

$$y = A - B = \sqrt[3]{\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} - \sqrt[3]{-\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}},$$

which now allows him to solve the original cubic by substituting the above values for p, q, and using $x = y - \frac{a}{3}$.

Somewhat ironically, this has become known as *Cardano's formula* (although some authors remember to credit del Ferro and/or Tartaglia). From Cardano's perspective, moreover, the formula only deals with one of the 'cases' he considered, namely $y^3 + py = q$, since he wishes to avoid negative coefficients at any stage.

This means (for example) that, in order to deal with the case $y^3 = py + q$ for positive p, q, he cannot simply write $y^3 - py = q$ and use the above argument with $-p$ instead to arrive at the solution! Instead he makes an elaborate substitution, based on the identity $(A+B)^3 = A^3 + B^3 + 3AB(A+B)$,which he justifies geometrically. We will soon discover why in such cases his formula would cause him major conceptual difficulties of a different kind.

As noted above, a general solution method for the general *quartic* (fourth-degree) equation of the form

$$x^4 + ax^3 + bx^2 + cx + d = 0,$$

found by *Ferrari* in 1540, was also included in Cardano's *Ars Magna.*

In brief, this solution method proceeds via the substitution $y = x - \frac{1}{4}$ to turn the general quartic equation into the form $y^4 + py^2 = -qy - r$, for appropriate new coefficients p, q, r which can again be expressed in terms of a, b, c, d. To attack the reduced equation, Ferrari notes that in the perfect square $(y^2 + \frac{p}{2})^2 = y^4 + py^2 + \frac{p^2}{4}$, the first two terms are what we have on the left in the above, so that $(y^2 + \frac{p}{2})^2 = \frac{p^2}{4} - qy - r$. He now adds a further unknown z to $(y^2 + \frac{p}{2})$ on the left, and again computes the square, substituting $(y^2 + \frac{p}{2})^2$ as above:

$$((y^2 + \frac{p}{2}) + z)^2 = 2zy^2 - qy + (z^2 + pz + \frac{p^2}{4} - r)$$

On the left we have expressed this quantity as a quadratic in y, and as it is a perfect square, so we need to find the value of z such that the discriminant of this quadratic is 0.

Recall that the *discriminant* of the general quadratic $ax^2 + bx + c = 0$ is $b^2 - 4ac$. In the present case we have $a = 2z$, $b = -q$, $c = (z^2 + pz + \frac{p^2}{4} - r)$, so that

$$q^2 = 4(2z)(z^2 + pz + \frac{p^2}{4} - r).$$

This cubic equation in z can be solved by use of Cardano's formula. Let $z = z_0$ be such a solution. For this value of z, the above quadratic in y has a double root $y_0 = \frac{q}{4z_0}$ (since the discriminant is zero, the root is $\frac{-b}{2a}$ in general). But now our equation in y, z above has the form $((y^2 + \frac{p}{2}) + z)^2 = 2z_o(y - y_0)^2$. So, for these values, the reduced quartic splits into a pair of quadratics whose solutions form the roots of the equation[a].

[a]This outline is based on: Ferrari method. Encyclopedia of Mathematics. URL: http://www.encyclopediaofmath.org/index.php?title=Ferrari_method&oldid=35675

And there, despite continuing efforts, matters would rest for more than 250 years. No-one was able to produce a formula similar to the ones described above in order to find the roots of polynomials of degree 5 (quintics) or higher. Nonetheless there was significant, if gradual, progress in the general understanding of the structure and theory of polynomial equations throughout the 17th and 18th centuries.

In 1799 the youthful German mathematician *Carl Friedrich Gauss* (1777-1855) asserted, without giving a proof, that the quintic has *no general solution formula* by means of radicals (i.e. using roots as was done above). Between 1799 and 1813 the Italian medical doctor, philosopher and mathematician *Paolo Ruffini* (1765-1822) published six versions of what he maintained was a proof that polynomial equations in powers higher than 4 have no such solution formulae, although his proofs were opaque and all contained significant flaws.

The brilliant young Norwegian mathematician *Niels Abel* (1802-1829), read Ruffini's work as a student and recognised that it was incomplete. However, by 1824 he had found a rigorous proof of Gauss' claim: there is no single *general formula* which will yield the solution of every quintic. Abel travelled to Paris, then very much the leading centre for algebra, hoping to develop his methods further by collaborating with the leading French mathematicians of the time. However, he found them unresponsive and soon labelled them as *'monstrous egotists'*, unwilling to collaborate with each other and especially with foreigners!

The brilliance of Abel's mathematical insights was not recognised during his short life. He suffered poverty and ill health, unable to find a university position. Just as others began to appreciate the outstanding merit of his work, he died of tuberculosis, aged 26, having by then turned away from polynomial equations. His work in other fields, especially in what are now known as elliptic and 'abelian' functions, was far-sighted and led to major advances in both number theory and algebra.

Three years later, modern algebra was revolutionised by the tragic, irascible French prodigy *Èvariste Galois* (1811-1832), who, as a left-wing firebrand, was mortally wounded at the tender age of 21 in a duel. Galois'

methods explained, as a by-product of a much broader algebraic theory now named after him, why the general quintic and higher-order polynomial equations can have no such solution formula. The technical details are well beyond the scope of this book: a description of these events is given in **[35]**.

3.2. Imaginary roots. The renown of the work of Cardano and his compatriots led to much greater interest in algebraic techniques from the late sixteenth century onwards. Cardano continued to describe negative numbers as 'fictitious' (*numeri ficti*) and ignored them when they occurred in his formula. He also encountered difficulties when solving equations of the form $y^3 = py+q$, since in that case the quantity under the square root sign in Cardano's formula is $(\frac{q}{2})^2+(\frac{-p}{3})^3$, which can become negative for particular values of p and q.

Consider, for example, the cubic equation $y^3 = 15y + 4$, which clearly has the solution $y = 4$. Cardano's formula for this equation yields

$$y = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

Cardano could not deal with the term $\sqrt{-121}$, since square roots of negative numbers seemed to make even less sense than negative numbers themselves! He could not understand why his formula would not yield the obvious root $y = 4$. He had noticed that a similar problem arises with quadratic equations: in examples of the form $x^2 + b = ax$ he realised that the usual solution formula would involve the square root of a negative number if $a^2 < 4b$. An example of this kind, included in his later work *Ars magnae sive de regulis algebraicis liber unus*, has $a = 10, b = 40$. He writes this equation as $x(10 - x) = 40$. He proposes $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$ as solutions, but still regards these as *'impossible'*, while conceeding that they are *'operationally'* correct. He regards $\sqrt{-15}$ as meaningless, calling it a *'quantita sophistica'*.

Two decades later, *Rafael Bombelli* (1526-ca.1572), aware that the cubic equation $y^3 = 15y + 4$, which had so troubled Cardano, has the solution 4, had what he called a *'wild thought'*. Perhaps, he argued, one could make sense of 'numbers' of the form $a + b\sqrt{-1}$ (for $2 + \sqrt{-121}$ take $a = 2$ and $b = 11$) by setting out *multiplication tables* for $\sqrt{-1}$, similar to those already in use for positive and negative numbers – although even these had not yet been properly justified! By analogy with what he knew from working with $+$ and $-$ for integers, including, notably $(-1)(1) = -1$ and $(-1)(-1) = 1$, he argued that the square root of a negative number *'has different arithmetical operations from the others'*.

These analogies led him to propose new multiplication rules that we would today write as:

$$(\sqrt{-1})(\sqrt{-1}) = -1 = (-\sqrt{-1})(-\sqrt{-1})$$
$$(-\sqrt{-1})(\sqrt{-1}) = 1 = (\sqrt{-1})(-\sqrt{-1}).$$

Bombelli articulated these relationships by means of verbal expressions, using *piu di meno* for $\sqrt{-1}$ and *meno di meno* for $-\sqrt{-1}$. For example, he expressed his key new rule (which we write as $(\sqrt{-1})(\sqrt{-1}) = -1$) as *piu di meno via piu di meno fa meno.*

Having experimented with such 'multiplication tables' for the square roots of *numeri ficti*, he proceeded to apply these in solving the troublesome equation $y^3 = 15y + 4$. Taking the cube of each term in the Cardano formula $y = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$, he obtained the (verbal) equivalent of

$$2 + \sqrt{-121} = (a + b\sqrt{-1})^3$$
$$2 - \sqrt{-121} = (a - b\sqrt{-1})^3$$

for some unknowns a, b. Multiplying out the perfect cubes on the right and using his new multiplication tables for $\sqrt{-1}$, he found (see *MM* for the calculation) that the choices $a = 2, b = 1$ provide a solution. Using this in the Cardano formula, Bombelli obtained the positive integer solution $y = (2 + \sqrt{-1}) + (2 - \sqrt{-1}) = 4$ for the equation! Thus, Bombelli's multiplication tables, used in Cardano's formula, produced the positive solution that was apparent by direct inspection, even though the terms of the formula appeared to include quantities that Cardano had considered to be 'pure fictions'.

Although he did not claim that the square root of a negative number should be accepted as a *number,* Bombelli had shown that using his rules for the multiplication of such quantities gave the correct positive root of this troublesome cubic equation. Bombelli was not able to make logical sense of his discovery, but he had in fact derived the correct multiplication rules for working with $\sqrt{-1}$, which we today call the *imaginary unit* i in the complex plane, to be discussed in **Chapter 4.** His calculation, while remaining speculative, did not banish the suspicion that such quantities had aroused, but perhaps it served to diminish it somewhat.