CHAPTER 1

NUMERALS AND NOTATION

FOR CENTURIES MATHEMATICS WAS DEFINED as "the science of number and magnitude", and while this definition cannot nowadays be taken as adequate, it does, nevertheless, reflect the origins of mathematics with reasonable fidelity. Notions related to the concepts of number and magnitude can be traced back to the dawn of the human race. Indeed some animal species—whose origins antedate those of humanity by millions of years—behave in such a way as to reveal a rudimentary mathematical sense: experiments with crows, for example, have shown that birds can distinguish among sets containing up to four elements. At any rate, mathematical thinking has long played a role in the everyday practical life of human beings.

The origin of the *number concept*, in particular, would seem to lie in our remote ancestors' grasping the idea of *plurality*, and seeing that pluralities or collections of things can be both *matched* and *compared in size*. For example, the hands can be matched with the feet or the eyes, the fingers with the toes, but not the feet with the fingers. The realization that matchability of hands, feet, eyes, and any other pair of objects is independent of their nature must have provided a crucial stimulus for the emergence of the idea of number. That small numbers such as two and four played an early and important role in human thought is shown by their special position in the grammar of certain languages, for example Greek, in which a distinction is made between one, two, and more than two, and Russian, which uses one noun case with numbers up to four, and a different one for larger numbers.

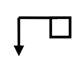
Numbers are assigned to collections by means of the process of *counting*, that is, the procedure of matching the elements of a collection successively with the ascending sequence of numbers, or number names. The recognition that the procedure of counting "one, two, three, four, ..." can be performed *intransitively*, in other words, that when counting it is not necessary to be actually counting *something*, is likely to have been instrumental in establishing the universality of the number concept. Indeed, it has been suggested that the art of counting arose in connection with primitive religious ritual and that the counting or *ordinal* aspect of number preceded the emergence of the quantitative or *cardinal* aspect. Whatever its origin, the procedure of counting naturally imposes an *order* on numbers, and it must have been grasped very early on that this order corresponds faithfully to the relative sizes of the collections that numbers are used to count.

As the awareness of number developed, it became necessary to express or represent the idea by means of *signs*. The earliest and most immediate mode of representing numbers is by means of the fingers, which conveniently enable numbers up to ten to be represented. For larger numbers heaps of pebbles were used, piled in groups of five, each corresponding to the number of fingers on the hand. Indeed our very term "calculate" derives from the Latin word *calculus* meaning "small stone". Counting by fives and tens—and, in some cultures, twenties—became standard practice, largely displacing the earlier systems of counting by twos and threes. Thus arose the idea of a *base* or *scale* for counting, in which some number b is selected, names for (some of the) numbers $1, \dots, b$ are assigned, and names for numbers larger than b are formulated as combinations of these. It was observed by Aristotle that the customary choice of base 10 is merely the result of the accidental fact that human beings happen to possess five fingers on each of two hands: from a strictly mathematical point of view it is somewhat regrettable that our ancestors did not possess a composite number of fingers, such as four or six, on each hand, rather than the awkward prime number five.

Heaps of pebbles, lacking ready portability as well as the requisite stability for prolonged storage of numerical information, came to be replaced in prehistory by less ephemeral and more portable devices, such as notches in a piece of wood or bone. The oldest known example of such a device is a tally stick dating from paleolithic times found in Moravia in 1937. It is the radius bone of a young wolf on which are incised 55 notches, arranged in two series, with 25 in the first and 30 in the second, both assembled in groups of five. The existence of these and a few other, similar artifacts, being more than thirty thousand years old, show that the use of the number concept, and, in particular, that of the number five, long antedates such technological advances as metal smelting or wheeled vehicles, and even the development of written language. This latter fact suggests that the appearance of number *signs* preceded that of number *words*, a plausible claim in any case since primitive signs for numbers such as notches or strokes possess a visual immediacy which has no counterpart in speech. The primacy of the *sign* or *symbol* in mathematics has persisted until the present day.

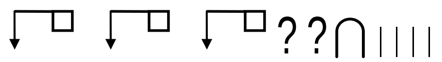
The earliest *numerals* or formal signs for numbers of which we possess definite record appeared in Egypt around 3400 B.C. The Egyptian system of *hieroglyphic* numerals employed strokes for numbers below ten, and special symbols for powers of ten, for example:

4 |||| 10 ∩ (a heel bone) 12 ||∩ 20 ∩∩

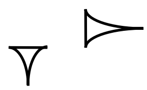
100 ? (a scroll) 200 ?? 1000  (a lotus flower)

There was even a symbol for 10000000 in the form of an astonished man or a god supporting the universe.

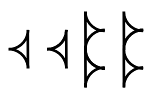
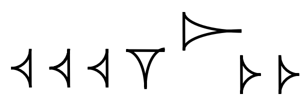
Numerals appeared in Egypt long before the first known numerical inscriptions in India (3rd century B.C.), China (3rd century B.C.) and Crete (1200 B.C.). Egyptian numerals constitute what is known as a *simple grouping system*. In such systems some number b—called a *base* or *scale* (10 in the Egyptian, as in most other cases)—is chosen and the powers b, b^2, b^3, ... of b treated as units—the *higher units.* Symbols for the latter, as well as for 1, are introduced and arbitrary numbers then represented by deploying these symbols *additively*, each symbol being repeated the requisite number of times. Thus, for example, in Egyptian hieroglyphics written left to right[1],

3215 = 

Another example of a simple grouping system is furnished by the Babylonian (c.2000 B.C.) symbols for numbers below 60. The early Babylonians (Sumerians) used clay as a medium for writing, producing inscriptions—the so-called *cuneiform* characters—by means of a stylus with a wedge-shaped tip. On the resulting clay tablets numbers less than 60 were expressed by a simple grouping system to base 10. Here we find a novelty, for the inscriptions are often simplified by the use of a *subtractive* symbol, viz.,



The numbers 1 and 10 were symbolized by  and . Thus, for example,

24 =  28 = 30 – 2 = 

As a further example of a simple grouping system we may consider the *Attic Greek* numerals developed some time prior to the 3rd century B.C. Here the powers of 10 are symbolized by the initial letters of number names; in addition there is a symbol for 5. These initial numerals, in modern notation, were

Π (Γ), *pi*, for ΠENTE (pente) five
Δ , *delta*, for ΔEKA (deka), ten
H, an old breathing symbol, for HEKATON, hundred
X, *chi*, for XIΛIOI (chilioi), thousand
M, *mu,* for MYPIOI (murioi), ten thousand

Thus, for example, they wrote

ΓΔ pente-deka, for 5 × 10 = 50

ΓH pente-hekaton, for 5 × 100 = 500,

and so

[1] As in Semitic languages, the Egyptians generally wrote from right to left.

3756 = XXX⌈HHH⌈Δ Γ| .

By far the most familiar simple grouping system is of course that based on the *Roman numerals.* Although nowadays it has only an ornamental use, the Roman system had the merit that the majority of its users needed to commit to memory the values of just four letters: V (5), X (10), L (50) and C (100). For larger numbers the symbols D (500) were employed. The simple grouping of symbols in the Roman system was eventually combined with the *subtractive principle*, in which a symbol for a smaller unit placed before a symbol for a larger one indicates that it is to be subtracted from the latter, as, e.g.

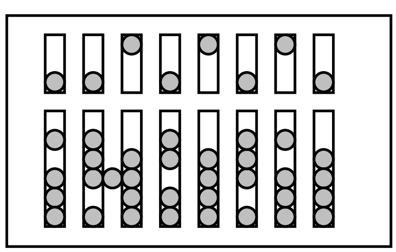
1949 = MCMXLIX.

The subtractive principle was employed only sparingly in ancient and medieval times, when the above number would have been written less succinctly

1949 = MDCCCCXXXXVIIII.

The question of the origin of the form of the Roman numerals has aroused considerable speculation. Of the various theories it is that of the German historian *Theodor Mommsen* (1817-1903) which has the strongest epigraphic support, and has gained the widest acceptance by scholars. Mommsen thought that the use of V for 5 resulted from its similarity in shape to the outspread hand with its 5 fingers, and that two of these gave the X for 10. He also contended that three of the other numerals were modified forms of Greek letters not employed in the Etruscan and early Latin alphabet. These were *chi*, which appears in inscriptions not only as **X**, but also in such forms as ⊥, ↓, which were to become the symbol L which was arbitrarily chosen to denote 50; *theta*, Θ, chosen to represent 100, and which, through the influence of the word *centum* (100), finally became transformed into C; and *phi*, Φ, to which was assigned the value 1000, and whose written form (|) gradually became, under the influence of the word *mille* (1000), the symbol M. With delightful literalness, half of (|) , i.e., |) , was taken to represent 500; this in turn metamorphosed into the later Latin D.

Roman numerals were poorly adapted for use in calculation, and actual computations were carried out on an abacus (from Greek *abax*, “slab”). Originally taking the form of sand or dust covered boards, by the late Roman period these devices had evolved into bronze tablets containing a number of grooves in which fixed counters slide. In each column a lower counter represents one, and the upper counter five, higher units.



Multiplicative grouping systems are refinements of simple grouping systems. In a multiplicative grouping system, the selection of a base b and symbols for 1 and the higher units b, b^2, b^3,... is augmented by a set of symbols for 2, 3, ..., $b-1$: these latter, together with 1 are known as *digits*. The symbols of the two sets are then employed *multiplicatively* to indicate the number of higher units required to form a given number. Thus, if we designate the first nine numbers by the usual symbols 1, ..., 9 but agree to designate 10,100,1000 by u, v, w, say, then in the corresponding multiplicative system we would write, for example

$$7843 = 7w8v4u3.$$

The traditional Chinese numeral system, usually written vertically, is an example of such a system to base 10.

Although concrete historical evidence is lacking, it may be surmised that at some point users of multiplicative grouping systems came to see that the values of the higher units could be indicated *locally*, that is, by the mere *position* of the digits, thereby enabling the symbols for higher units to be dispensed with altogether. But in so doing ambiguities will arise unless a new digit—a *zero*—is introduced to signalize powers of the base that happen to be missing in the representation of a given number. (For instance, without such a symbol we would be unable to distinguish, say, 22 from 202.) Thus, in a *positional* number system on a base b, the symbols for the higher units are discarded and the digits augmented by a symbol for zero; each number N can then be uniquely expressed as a sum

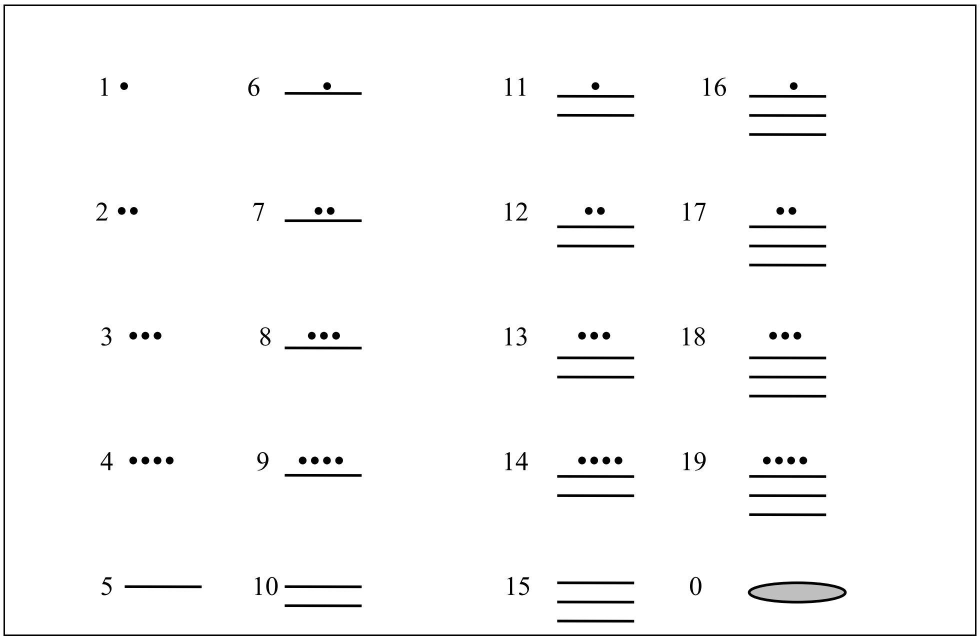
$$N = a_n b^n + a_{n-1} b^{n-1} + \ldots + a_2 b^2 + a_1 b + a_0 ,$$

where $a_0, \ldots, a_n$ are digits, so enabling N to be presented in the succinct form of a sequence of digits

$$N = a_n a_{n-1} \ldots a_2 a_1 a_0.$$

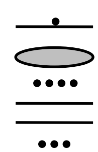
The earliest number system incorporating the positional principle was developed by the Babylonians before 2000 B.C.: here numbers exceeding 60 were written *sexagesimally*, that is, positionally in base 60. Lacking a symbol for zero, however, the system inevitably suffered from ambiguity. By 200 B.C. the Babylonians' successors had introduced a symbol—a proto-zero, so to speak—to indicate the absence of a number, but did not employ it in calculation, showing that they had not yet taken the crucial step of ascribing it full digithood.

The first fully positional system of enumeration was developed by the Mayas of Yucutan around 300 A.D. Scholars have come to believe that the large numbers inscribed on their monuments and in their few surviving bark-paper codices are the dates of lunar and solar eclipses presented in terms of the "Long Count", a cyclic calendar of religious origin apparently dating back many centuries. The Maya system is remarkable in its employment of a *mixed base* in addition to a symbol for zero: although essentially *vigesimal*, that is, based on 20, the third digit position is $18.20 = 360$ instead of $20^2 = 400$ and subsequent digit positions are of the form 18.20^n. This curiosity almost certainly reflects the calendric origins of their numeration system, for the official Maya year had 360 days. The Maya employed just three symbols: a dot for the unit, a bar for five, and a stylized shell for zero. Numbers below twenty were notated according to a simple grouping scheme



Larger numbers were written vertically, for example:

$$43483 = 6.18.20^2 + 0.18.20 + 14.20 + 3 \quad =$$



The system of numeration used almost universally today is the *Hindu-Arabic decimal system*, believed by scholars to have been introduced in India sometime before the ninth century A.D. and subsequently developed by Arab scholars who later transmitted it to the West. Not the least of its advantages is the fact that it enables calculations to be performed simply, in a purely symbolic manner, rendering unnecessary the use of an auxiliary device such as an abacus. In fact, the obvious correspondence between the lines of an abacus and the digit positions in the Hindu-Arabic representation of numbers suggests that the system may have originated as a symbolic representation of the abacus itself: this would make natural the introduction of a symbol for zero as corresponding to a line on the abacus empty of counters. In any event, by the sixteenth century the "algorists"—champions of the Hindu-Arabic system—had achieved supremacy in Europe over their rivals the "abacists"—those who still employed Roman numerals for the recording of numbers and the abacus for computation[2].

Once the Hindu-Arabic system had been fully accepted other advantages came to light, for instance, the fact that it provides a convenient notation for *fractions*, using inverse powers of 10. For example, 3/8 is represented by the *decimal*

$$0.375 = 3 \times 10^{-1} + 7 \times 10^{-2} + 5 \times 10^{-3}.$$

Of course, in this system some fractions will be represented by nonterminating (periodic) decimals, e.g.

$$\tfrac{1}{3} = 0.33333.... = 3 \times 10^{-1} + 3 \times 10^{-2} + 3 \times 10^{-3} + ...,$$

in which the sum on the right approaches $\tfrac{1}{3}$ arbitrarily closely as more terms are "added in". Expressing fractions in decimal notation enables calculations with them to be performed in exactly the same way as with whole numbers.

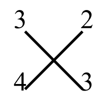
The actual form of the Hindu-Arabic numerals underwent considerable variation before being finally stabilized by the invention of printing in the fifteenth century. It has been suggested that the numerals are of *iconic* origin: specifically, that they were originally formed by drawing as many strokes or dots as there are units represented by the respective numerals, these strokes or dots later becoming connected through the use of cursive writing. While this theory does not seem entirely implausible (at least in the case of 1, 2, and 3), it is not supported by solid historical evidence, nor in any case does it explain the great variety of forms which the numerals took at different times and places.

The symbol "0" for zero may have evolved from the dot "·" which was first used by the Hindus. Before definitively assuming its present form, the symbol "0" was sometimes crossed by a horizontal or vertical line, causing it to resemble the Greek letters "Θ" or "Φ". The word "zero" itself seems to have come through Old Spanish from the Arabic *sifr*, "empty", from which our word *cipher* also derives.

Greater confidence may be placed in theories accounting for the form of the symbols for *operations* on numbers, such as addition and multiplication, since these are of considerably more recent origin than the signs for the numbers themselves. The modern algebraic symbols "+" and "–" for addition and subtraction first appear in fifteenth-century German manuscripts. It is highly likely that the "+" sign descended from a cursive form of *et*, "and", in Latin manuscripts. Concerning the origin of the minus symbol "–" nothing certain is known, but it may simply have arisen as a sign indicating separation. The sign "×" (St. Andrew's cross) for multiplication first appears in *William Oughtred's* (1574–1660) *Clavis Mathematicae* of 1631, but it seems likely

[2] It is an irony that the development of the electronic computer —the abacus of today—is rapidly reducing the role of the Hindu-Arabic numeral system to that of its Roman counterpart, i.e., to a device for the mere recording of numbers. But this "abacists' revenge" has come at a price, since, unlike the lines of an abacus, the internal structure of a computer is far too complex to suggest the form of *its* immediate successor.

that he derived it from earlier uses of the cross in solving problems involving proportion. Thus, for example,



was used to indicate that 3 was to be multiplied by 3, and 4 by 2. The use of the dot "·" for multiplication[3] was explicitly introduced by the German mathematician and philosopher *Gottfried Wilhelm Leibniz* (1646–1716) in 1698 in a letter to John Bernoulli, in which he states: "I do not like × as a symbol for multiplication as it is easily confounded with *x*… often I simply relate two quantities by an interposed dot and indicate multiplication by ZC · LM." The sign "÷" for division was originally employed for subtraction; its first recorded use as a sign for division (deriving perhaps from the fractional line "—") is in a Swiss Algebra of 1659. Its use then quickly spread to the English speaking countries, but not to the rest of continental Europe, where Leibniz's symbol ":" for division remained in force well into the twentieth century.

The familiar sign "=" for equality was introduced by the English mathematician *Robert Recorde* (c. 1510–1558) in his charmingly titled algebra treatise *The Whetstone of Witte*, published in 1557. Recorde says he was led to adopt a pair of equal parallel line segments as the symbol for equality "bicause noe 2 thynges can be more equalle." Despite its simplicity and suggestiveness, Recorde's sign did not gain immediate acceptance; in fact for two centuries a struggle for supremacy took place between it and *René Descartes'* (1596–1650) sign "⋊" — possibly derived from the first two letters of the Latin *aequalis*, "equal"— which he had introduced in his *Géométrie* of 1637. The final victory of = at the close of the seventeenth century seems to have been mainly due to the influence of Leibniz, who had adopted Recorde's symbol.

The signs ">" and "<" for "greater" and "less" were introduced by the English mathematician *Thomas Harriot* (1560–1621) in his work *Artis Analyticae Praxis* of 1631. Like Recorde's sign for equality, the use of these signs—despite their simplicity and suggestiveness—spread surprisingly slowly, not becoming standard before the nineteenth century.

[3] It is interesting that in algebra multiplication is indicated by *juxtaposition*: thus *ab* stands for "*a* times *b*." This may derive from the fact that in everyday language number terms can be used adjectivally, as in "four apples"; when the modified noun is itself a number, the resulting phrase denotes a multiplication: for example, "four threes" is taken to mean "four times three", i.e. 12. And indeed the multiplication tables were (until recently) learned precisely by repeating phrases of this sort. Juxtaposition is of course systematically employed in positional number systems to indicate *addition* of numerals, and this probably explains why the same device is employed to indicate the addiing of a fraction to an integer: thus, for example, "5" stands for "5 + ¾".