Chapter 1: Whole Numbers (Draft)

H. Wu

Department of Mathematics #3840
Universit of California, Berkeley
Berkeley, CA 94720-3840
http://www.math.berkeley.edu/∼wu/

wu@math.berkeley.edu

Contents

[0] **September 1, 2002**

How to Read This Monograph

In order to learn mathematics, you will have to read articles and books in mathematics, such as this monograph. Reading mathematics requires a different skill from reading novels or magazines, and I want to say a few words about this difference.

First of all, reading mathematics requires sustained effort and total concentration. It is a slow and painstaking process. This monograph is not a great candidate for bedtime reading unless insomnia is not a problem with you. Some mathematics textbooks, especially those written in the last three decades, do a lot of "padding", i.e., inserting long passages with little content, and may have instilled the illusion that skipping is a good policy in the reading of mathematics. This monograph, by contrast, says only what needs to be said, so you will have to read every line AND *try to understand every line. In fact, often you will find yourself struggling to understand every word in order to move forward. On occasion, I do some chatting, but more often than not I will be talking about straight mathematics. I have made every effort to supply you with sufficient details to follow the reasoning with ease, but I tend not to waste too many words. So you would have to read everything carefully. In the event that I believe something can be safely skipped on first reading, I leave it in* INDENTED FINE-PRINT PARAGRAPHS.

You may have gotten used to the idea that in a mathematics book you only need to look for soundbites: understand a few procedures and forget the rest. Not so here. This monograph tells a coherent story, but the outline of the plot (the PROCEDURES*) is already familiar to you. It is the details in the unfolding of the story (the* REASONING*) that are the focus of attention here. Think of yourself as a detective who has to solve a murder case: you already know going in that someone was killed, so you job is not just to report the murder but to to find out who did it, how he did it, and why he did it. It is the details that matter, and they matter a lot. Learning the details of anything is hard work. I want to to tell tell you that most mathematicians also regard learning mathematics as very hard work. It takes time and effort, and it may mean being stuck for a long time trying to understand* *a particular passage. Nothing good comes easily.*

It would be futile, not to say impossible, for me to anticipate the kind of difficulties each of you may have in reading the monograph. Experience would seem to indicate, however, that most of you will be surprised by the emphasis in this monograph on the importance of DEFINITIONS. *A (very) mistaken belief which unfortunately has gained currency in recent years is that, in the same way that children learn to speak whole sentences without first finding out the precise meaning of individual words, students can also learn mathematics by bluffing their way through logical arguments and computations without finding out the precise meaning of each concept. As a result, it is customary in schools to teach mathematics using mathematical concepts that are only vaguely understood.* **Such a belief is completely without foundation.**

Take the concept of a fraction or a decimal, for example. It is almost never clearly defined. Yet children are asked to add, multiply and divide fractions and decimals without knowing what they are or what these operations mean, and textbooks contribute to children's misery by never defining them either. If we can get away with this kind of mathematics education, — in the sense that children learn all they need to learn without the benefit of clear definitions — fine. BUT WE CANNOT, *because children are on the whole not learning it. From the standpoint of mathematics, the first remedy that should be tried is to explain clearly what these concepts mean, because mathematics by its very nature is a subject where* EVERYTHING *is clearly explained. Giving clear definitions of concepts before putting them to use has the virtue of taking the guesswork out of learning: every step can now be explained, and therefore more easily learned. This is the approach taken here. If you feel uncomfortable with such an approach, can you perhaps suggest an alternative? In any case, it is only a matter of time, and maybe a little practice, before you get used to it. (Smokers also feel extremely uncomfortable at the beginning of their attempt to quit smoking.) You will discover that having clear-cut definitions is by far the better way to learn* AND *to teach mathematics.*

At the risk of stating the obvious, I may point out that while *this monograph addresses serious mathematics, its exposition is given in ordinary conversational English (or as conversational as an ESL person can manage it). Why this is worth mentioning is that there is at present a perception that mathematical writing should not be couched in ordinary English. The thinking goes roughly as follows. Because mathematics is somehow* DIFFERENT*, it requires a* DIFFERENT *kind of writing: fewer words and more symbols, for instance, and complete sentences are optional. Whatever the justification of this kind of misconception, the end result is there for all to see: students stop using correct gammar and syntax in their homework and exam papers, and a random collection of symbols out of context usually passes for an explanation. If we want to change such behavior among students, we would do well to first change this misconception about mathematical writing amongst ourselves. We should never forget that mathematics is an integral part of human culture. Doing mathematics is above all a normal part of human activities, and it imposes on us the same obligations of normal human communication as any other endeavor. We must make ourselves understood via the usual channels in the usual manner. The subject matter requires greater precison of expression than a chat about the private lives of movie stars, to be sure, but this precision is something we try to achieve* IN THE CONTEXT OF *normal communication, rather than* IN SPITE OF *it. Please keep this in mind as you read this monograph.*

Chapter Preview

This chapter discusses the *whole numbers*

$$0,\ 1,\ 2,\ 3,\ 4, \ldots$$

with a view towards laying a firm foundation for the treatment of the main topics of this monograph, namely, fractions and decimals. Notice that we include 0 among the whole numbers. The main emphasis throughout will not be on the well-known procedures such as the long division algorithm — although a precise and correct statement of that algorithm is certainly difficult to find in the literature — but on the logical reasoning that underlies these procedures. In mathematics, be it elementary or advanced, the first question you should always ask when confronted with any statement is "Why?". To try to find out why something is true is a very natural human impulse. Should you have any doubts, just observe how often pre-school children raise this simple question with their parents each time they are introduced to something new. As a teacher, your obligation is to keep alive this sense of curiosity in a child. One way to do this is to ask yourself the same question at all times *and* to find out the answers, because it is also your obligation to answer this question for your students. To this end, this chapter will revisit a very familiar territory, — the arithmetic of whole numbers — with the goal of explaining everything along the way. Because everything here *is* familiar to you, at least as far as procedures are concerned, there is an inherent danger that as you read this material you would put yourself on automatic intellectual pilot and cease to think. To get you out of this counter-productive mode, I would explicitly ask you to put yourself in the position of a *first-time learner* and to make believe that you are encountering every topic for the first time. I realize that it is very difficult to do this because it requires a suspension of habits. Nevertheless, learning this material is so crucial for the understanding of the rest of the monograph that I must ask you to please try your best. Once you get the hang of it, you would not only acquire an enhanced appreciation of the marvelous qualities of many things you have always taken for granted, but also find yourself in a much better position to learn the materials in the later chapters on decimals and fractions.

1 Place Value

One cannot understand the arithmetic of whole numbers without a basic understanding of our numeral system, the so-called *Hindu-Arabic numeral system.*[1] It became the universal numeral system in the West *circa* 1600. This chapter discusses only the whole numbers 0, 1, 2, 3, . . . , and the discussion of this numeral system will continue in Chapters 4 and 5. We are here mainly concerned with the fact that a symbol such as 2 in the number 2541 stands not for 2 but 2000. In fact, 2541 means

$$2000 + 500 + 40 + 1 \ ,$$

i.e., two thousand five hundred and forty-one. As is well-known, the ten symbols 0, 1, 2, . . . , 9 are called *digits*. The digit "2" in 2541, being in the fourth *place* (position) from the right, stands not for 2 but 2000, i.e., two *thousand*, as mentioned above. Similarly, the digit "5" being in the third place from the right stands not for 5, but for 500, i.e., five *hundred*, the "4" in the second place from the right stands for 40, and "1" being in the right-most place means just 1. Similarly, 64738 denotes

$$60000 + 4000 + 700 + 30 + 8$$

and 6001 denotes

$$6000 + 1,$$

and so on. These examples illustrate a fundamental and fruitful idea of representing numbers, no matter how large, by the use of only ten symbols 0, 1, 2, . . . , 8, 9, and by the use of their place in the number symbol to represent different magnitudes (sizes). Thus the "5" in 125 represents a completely different order of magnitude from the "5" in 2541, namely, 5 and 500 respectively. The term *place value* means that the value (magnitude, size) of each digit depends on its place in the numeral symbol.

[1] This is most likely a misnomer. The Chinese had a decimal numeral system since the first available record of writing dating back to at least 1000 B.C., and the *rod numeral system* (also called the *counting board numeral system*) which has been firmly in place no later than 200 A.D. is *identical* to the Hindu-Arabic system except for the ten symbols 0, 1, 2, . . . themselves. Moreover, negative numbers and decimal fractions (see Chapter 4) have been part of the rod numeral system from the beginning. Because of the long history of contact between the Indians and the Chinese, it may be difficult to separate what is Indian and what is Chinese in the Hindu numeral system. The much needed research has not yet been done.

In case the idea of place value has become too commonplace to strike you as noteworthy, let us look at a different numeral system for comparison: in Roman numerals,[2] the number 33 is represented by XXXIII. Observe then that the three "X's" are in three different places, yet each and every one of them stands for 10, not 100 or 1000. Just 10. Similarly, the three "I's" occupy different places too, but they all stand for 1, period. Contrast this with the numeral 111 in our numeral system: the first 1 on the left stands for 100, the second stands for 10, and only the third stands for 1 itself. You see the difference.

We have used the concept of addition to explain place value (e.g., $125 = 100+20+5$). We could have pretended that you didn't know what it means to add whole numbers and give a precise definition, but that would be too pedantic. People seem to have no problem with understanding this concept. But we will carefully and precisely define the other three arithmetic operations in this and the the next two sections, i.e., subtraction, multiplication, and division.

So far we have not brought out the significance of the fact that only ten symbols 0, 1, 2, …, 8, 9 — instead of twenty or sixty, say — are used to denote any number, no matter how large. We now fill in this gap. Like place value, the fact that only ten symbols are used is easy to overlook due to constant usage. It should be pointed out, therefore, that the great virtue of the Hindu-Arabic system lies precisely in the systematic and combined application of *both* ideas — place value and a fixed small number of symbols (ten, to be exact) — to generate all the numbers. No other numeral system of the world (except the rod numeral system of China, see footnote 1) has ever attained the same degree of symbolic economy. The Babylonians in the B.C. era, for instance, used a numeral system that used place value only partially, and the symbolic representation of some numbers became unwieldy. Since we have already mentioned the Roman numerals, let us use it to illustrate how complications arise from trying to cope with large numbers when place value is not systematically applied. The symbol with the largest numerical value in the Roman system is M, which denotes a thousand, 1000. In order to write a million, which is a thousand thousand, one would have to write

[2] It is well to point out that even in the Roman numeral system, there is a *partial* place value at work. For instance "VI" is 6 while "IV" is 4.

$MMM \dots MM$ (a thousand times). The Romans were spared this drudgery apparently because they never had to deal with a large number of this size. A latter day ad hoc convention to improve on the Roman system is to add a bar above each symbol to increase its value a thousand times. Thus $\overline{X}$ would denote not 10 but 10000, and $\overline{M}$ would denote a million. Even such a desperate effort cannot save this numeral system from certain disaster, however, because it would still be too clumsy. For instance, the simplest way to write 388999 is

$$\overline{CCCLXXXVIII}CMXCIX.$$

It is conceivable that had Roman numerals been adopted as the universal numeral system in the modern era, someone would have tried to introduce symbols for "ten thousand" and "hundred thousand" to simplify the writing of 388999, but if so, what about writing 60,845,279,037? Presumably, more symbols would yet be introduced. Would you want to waste your time learning how to navigate in such a system?

To truly understand place value, we must review the process of *counting* from 0, 1, 2, ... onward in order to see how the whole numbers develop in the Hindu-Arabic numeral system. After 0, 1, ..., 8, 9, we have used up all possible symbols by allowing ourselves only one place (position), the so-called *ones place*. To generate the next set of numbers without adding more symbols, the only option available is to put the same synbols in an additional place, the so-called *tens place*, which by convention is to the left of the ones place. (Keep in mind that this is no more than a convention.) In our minds, we may think of 0, 1, 2, ..., 8, 9 as 00, 01, 02, ..., 08, 09; in other words the single-digit numbers may be thought of as two-digit ones with 0 in the tens place.[3] From this point of view, the next number after 9 is naturally 10, i.e., since 0 has already been used in the tens place, we replace 0 by its successor 1. Thus we change the 0 of 09 to 1, and start the counting in the ones place all over again with 0. The numbers after 09 are then 10, 11, 12, ..., 18, 19. The same outlook then guides us to write the number after 19 as 20, because after having used up all the digits in the ones place with the tens place occupied by 1, it is natural to increase the latter from 1 to 2 and start the counting in the ones place all over again with 0. Thus the next numbers are 20, 21, 22, etc. Continuing this way, we get to 97, 98, 99. At

[3] There is a further discussion of this issue of having 0's to the left of a number at the end of this section.

this point, we have used up all ten digits in both the tens place and the ones place. To proceed further, and without introducing more symbols, we will have to make use of another place to the left of the tens place, the so-called *hundreds place* (the third place from the right). Thinking of 99 again as 099, the same consideration then dictates that the next number is 100, to be followed by 101, 102, etc. Thus we come to 109, and the next is 110, followed in succession by 111, 112, 113, ..., 198, 199. After that come 200, 201, 202, etc., for exactly the same reason.

We pause to note that, between the numbers 0 and 100, if we skip count by 10's, then we have 00, 10, 20, ..., 80, 90, 100. Thus in ten steps of 10's, we go from 0 to 100.

By the time we reach 999, again we have used up all ten digits in three places so that the next number will have to make use of a fourth place, the so-called *thousands place* (fourth place from the right). It will have four digits and it has to be 1000 because 999 can be thought of as 0999 and we naturally increase the 0 of 0999 to 1 and start counting all over again in the ones, tens, and hundreds places. In general, whenever we reach the number $99\ldots9$ (n times for any nonzero whole number n), the next number must be $100\ldots0$ (n zeros). As before, we make the following observation: if we count from 0 to $100\ldots0$ (n zeros) in steps of $100\ldots0$ ($n-1$ zeros) for any nonzero whole number n, then we have

$$0\underbrace{00\ldots0}_{n-1},\ 1\underbrace{00\ldots0}_{n-1},\ 2\underbrace{00\ldots0}_{n-1},\ \ldots\ ,9\underbrace{00\ldots0}_{n-1},\ 1\underbrace{000\ldots0}_{n}.$$

In other words, in ten steps of $100\ldots0$ ($n-1$ zeros), we get to $1000\ldots0$ (n zeros) from 0.

Thus we can make three observations about the way counting is done in the Hindu-Arabic numeral system: for any nonzero whole number n,

(i) *an n digit number precedes any number with more than n digits,*

(ii) *given two n digit numbers a and b, if the n-th digit (from the right) of a precedes the n-th digit of b, then a precedes b,* and

(iii) *the sum of ten* $1\underbrace{00\ldots0}_{n-1}$*'s is* $1\underbrace{00\ldots0}_{n}$.

In view of the previous comments about adding 0's in front of a number, we should add the following clarification: a number is said to be an *n digit* *number* if, counting from the right, the last nonzero digit is in the n-th place. For example, 0050000 is a 5-digit number, and 1234 is a 4-digit number. As illustrations of (i)–(iii): 987 precedes 1123, 65739 prcedes 70001, and

$$\underbrace{10000 + 10000 + \cdots + 10000}_{10} = 100000.$$

This is the right place to review and make precise the common notion of “bigger than”. Formally, for two whole numbers a and b, we define b to be *bigger than* a (or what is the same, a to be *smaller than* b) if, in the method of counting described above, a comes before b. In symbols:

$$a < b \qquad \text{or} \qquad b > a.$$

Note that one sometimes says *greater than* in place of *bigger than*, and *less than* in place of *smaller than*. If we want to allow for the possibility that b *is bigger than or equal to* a, then we write:

$$a \leq b \qquad \text{or} \qquad b \geq a.$$

Thus $13 \leq 13$ and $7 \geq 7$, but $7 < 13$ and $9356 < 11121$, etc. In particular, it is always the case that

if n is a nonzero whole number, then $n > 0$.

It follows from observations (i) and (ii) that

(iv) *if a, b are whole numbers and b has more digits than a, then* $a < b$, and

(v) *if a, b are two whole numbers with n digits and the n-th digit of a is smaller than the n-th digit of b, then $a < b$.*

For example, $872 < 1304$, $100002 > 99817$, $803429 < 911104$, etc.

We next turn our attention to the phenomenon of “too many zeros”. Consider a moderate-size number such as the number of seconds in a 365-day year: 31536000. As we know, this means:

$$30000000 + 1000000 + 500000 + 30000 + 6000.$$

I hope you are already weary of reading and trying to keep track of so many zeros. It could be much worse, of course. For example, the age of the universe has now been established (by observations of the Hubble telescope) to be approximately 14000000000 (9 zeros) years. Back in the third century B.C., Archimedes estimated that the number of grains that can be packed into a ball the size of the then-known universe was

$$1,000,000,000,000,000,000,000,000,000,000,000,000, \\ 000,000,000,000,000,000,000,000,000$$

(63 zeros). Clearly, we must devise a shorthand notation to deal with these zeros. (*In case you haven't noticed, symbolic notation has always arisen from the need for better human communication.* The Hindu-Arabic numeral system — as you have seen — is a good example.)

We need to introduce a new notation and to review an old one. First the new notation: we write 10^1 for 10, 10^2 for 100, 10^3 for 1000, 10^4 for 10000, 10^5 for 100000, and in general, write

$$10^n \qquad \text{for} \qquad 1\underbrace{00\ldots0}_{n} \tag{1}$$

where n is any whole number > 0. For notation consistency, we also write

$$10^0 \qquad \text{for} \qquad 1.$$

(You may think of 10^0 as "zero number of 0's".) The number n in 10^n is called the *power* or *exponent* of 10^n, and 10^n is read as "10 to the n-th power".

Next, we recall the definition of multiplication *among whole numbers* as a shorthand notation for repeated addition. In other words, 3×5 by definition $5 + 5 + 5$, and in general, if m, k are whole numbers, the definition of mk (the accepted abbreviation of $m \times k$) is:

$$mk = \begin{cases} 0 & \text{if } m = 0 \\ \underbrace{k + k + \cdots + k}_{m} & \text{if } m \neq 0 \end{cases} \tag{2}$$

Sometimes we refer to mk as the *product* of m and k, and call m and k the *factors* of mk. We call attention to the fact that mk, the multiplication of k by m, is **a shorthand notation** for adding k to itself m times, no more and no less. Please be sure to impress this fact on your students. (See Exercise 1.8.)

Activity: Consider the following introduction to *multiplication* taken from a third grade textbook (the text has the goal of making sure that at the end of the third grade, students know the multiplication table of numbers up to 10):

> Look at the 3 strips of stickers shown on the right (there is a picture of three strips of stickers). There are 5 stickers on each strip. How can you find the number of stickers there are in all?
>
> You can find the total number in different ways.

You can write an addition sentence. $5+5+5=15$ THINK: 3 groups of 5 = 15.	You can write a MULTIPLICATION sentence $3\times 5 = 15$ READ: Three times 5 equals 15.

> Answer: 15 stickers.

Do you think this is an ideal way to convey to third graders what *multiplication* means?

The definition of multiplication in (2) lends itself to an easy pictorial representation. For example, 3×5 , which is $5+5+5$, can be represented by three rows of five dots:

$$\begin{array}{ccccc} \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \end{array}$$

Similarly, mk $(= \overbrace{k+k+\cdots+k}^{m}\,)$ can be represented by m rows of k dots:

$$\underbrace{\begin{array}{cccc} \bullet & \bullet & \cdots & \bullet \\ \bullet & \bullet & \cdots & \bullet \\ \vdots & \vdots & & \vdots \\ \bullet & \bullet & \cdots & \bullet \end{array}}_{k}$$

There should be no mistaking the fact that the definition in (2) means *exactly* that mk is adding k to itself m times, and NOT adding m to itself k times. We are taking nothing about multiplication for granted, so that if we wish to say mk actually also equals adding m to itself k times, we would have to explain why. This will be done later in §2, but for now we don't need this distraction.

We now put the new information to use: if n is a nonzero whole number, then

$$\begin{aligned} 2 \times 10^n &= 1\underbrace{00\ldots0}_{n} + 1\underbrace{00\ldots0}_{n} = 2\underbrace{00\ldots0}_{n} \\ 3 \times 10^n &= 1\underbrace{00\ldots0}_{n} + 1\underbrace{00\ldots0}_{n} + 1\underbrace{00\ldots0}_{n} = 3\underbrace{00\ldots0}_{n} \end{aligned}$$

and in general, for $k = 1,\ 2,\ \ldots, 9$,

$$k \times 10^n = \underbrace{10^n + 10^n + \cdots + 10^n}_{k} = k\underbrace{00\ldots0}_{n}.$$

We can now revisit 31536000 and rewrite it as

$$31536000 = (3 \times 10^7) + (1 \times 10^6) + (5 \times 10^5) + (3 \times 10^4) + (6 \times 10^3).$$

(We recall the *convention concerning parentheses*: do the computations within the parentheses first.) Similarly, the age of the universe is approximately

$$14,000,000,000 = 10,000,000,000 + 4,000,000,000 = (1 \times 10^{10}) + (4 \times 10^9),$$

and Archimedes' number of grains of sand is simply 10^{63}.

In general, a whole number such as 830159 can now be written as

$$830159 = (8 \times 10^5) + (3 \times 10^4) + (1 \times 10^2) + (5 \times 10^1) + (9 \times 10^0)\,.$$

Such an expression is sometimes referred to as the *expanded form* of a number, in this case, 830159. Another example is

$$2070040 = (2 \times 10^6) + (7 \times 10^4) + (4 \times 10^1).$$

One advantage of writing a number in its expanded form is that it instantly reveals the true value of a given digit in the number. For example, the 3 of 830159 is given on the right as 3×10^4, which immediately signals that it stands for 30000 and not 3. The notation also gives a clear and precise location (place) of a digit in the number symbol: the exponent, say k, of 10 in the expanded form of a number indicates precisely that the associated digit is the $(k+1)$-th one from the right. Thus the expression 3×10^4 tells us that 3 is in the fifth place (from the right) of 830159, and 2×10^6 indicates the position of 2 as the seventh digit from the right of 2070040. The clear location of a digit in the expanded form of a number will turn out to be very helpful in understanding all the arithmetic algorithms in §3.

There is one aspect of the expanded form that may trouble you: why use the cumbersome notation of 5×10^1 and 9×10^0 in the expanded form of 830159 instead of just 5×10 and 9 ? The answer is: it all depends on what we want. When absolute conceptual clarity is called for, we will use 5×10^1 and 9×10^0, such as when we discuss the *complete expanded form of a number* later on in §3 of Chapter 4. On the other hand, the less rigid notation of 5×10 and 9 is sufficient for ordinary purposes, so usually we just write:

$$830159 = (8 \times 10^5) + (3 \times 10^4) + (1 \times 10^2) + (5 \times 10) + 9 \; .$$

It remains for us to tie up some loose ends in the foregoing discussion. The first one is that some of you may have encountered another definition of 10^n for a nonzero whole number n as

$$\underbrace{10 \times 10 \times \cdots \times 10}_{n},$$

whereas we have defined 10^n in (1) as

$$10^n = 1\,\underbrace{00\ldots0}_{n}\,.$$

We need to clarify this situation by proving that these two numbers are the same, i.e., we have to prove:

$$10^n = \underbrace{10 \times 10 \times \cdots \times 10}_{n} \qquad (3)$$

for all whole numbers $n > 0$.

Activity: Prove that (3) is true for $n = 1,\ 2,\ 3$.

You may think that (3) is obvious, but it could only be because you are used to thinking of n as a small number, say $n = 1,\ 2,\ 3$. But what about $n = 123457321009$? Can you even make sense of (3) in that case? Thinking about these questions will perhaps convince you that we should spend some time finding out why (3) is true for *all* values of n.

A second loose end we should tie up is a fact that many of you probably take for granted, namely, that for any number, e.g., 2,133,070, multiplying it by 100,000 (say) results in a number that is obtained from 2,133,070 by tagging the 5 zeros of 100,000 to the right of 2,133,070. That is,

$$2,133,070 \times 100,000 = 213,307,000,000$$

The last loose end we wish to address is the issue of the implicit zeros in front of any whole number, i.e., 0023 is the same as 23.

Let us first explain why (3) above is true. The reason is twofold:
(a) It is true for $n = 1$, and
(b) if we introduce the *temporary notation* that

$$10[n] = \underbrace{10 \times 10 \times \cdots \times 10}_{n},$$

then both symbols 10^n $(= 1\underbrace{00\ldots0}_{n})$ and $10[n]$ are "built up" in exactly the same way, in the sense that for any nonzero whole number n,

$$10^{n+1} = 10 \times 10^n \qquad \text{and} \qquad 10[n+1] = 10 \times 10[n]. \qquad (\odot)$$

Note that using the temporary notation, (3) may be restated as

$$10^n = 10[n] \qquad \text{for all whole numbers } n > 0.$$

Let us first make use of (a) and (b) to prove this before we prove (a) and (b) themselves.

For $n = 1$, it is quite obvious that both 10^1 and $10[1]$ are equal to 10 and therefore equal. Next, why is $10^2 = 10[2]$? This is because

$$\begin{aligned} 10^2 &= 10 \times 10^1 \qquad \text{(by first equality of } (\odot) \text{ with } n = 1) \\ &= 10 \times 10[1] \qquad (\text{by } 10^1 = 10[1]) \\ &= 10[2] \qquad \text{(by second equality of } (\odot) \text{ with } n = 1) \end{aligned}$$

Next we prove $10^3 = 10[3]$:

$$\begin{aligned} 10^3 &= 10 \times 10^2 \qquad \text{(by first equality of } (\odot) \text{ with } n = 2) \\ &= 10 \times 10[2] \qquad (\text{by } 10^2 = 10[2]) \\ &= 10[3] \qquad \text{(by second equality of } (\odot) \text{ with } n = 2) \end{aligned}$$

Next we prove $10^4 = 10[4]$:

$$\begin{aligned} 10^4 &= 10 \times 10^3 \qquad \text{(by first equality of } (\odot) \text{ with } n = 3) \\ &= 10 \times 10[3] \qquad (\text{by } 10^3 = 10[3]) \\ &= 10[4] \qquad \text{(by second equality of } (\odot) \text{ with } n = 3) \end{aligned}$$

At this point, it is clear that the proof of $10^5 = 10[5]$ will follow exactly the same pattern, and then $10^6 = 10[6]$, $10^7 = 10[7]$, etc. So we see that (3) must be true in general for all $n > 0$.

We now take care of (a) and (b). Clearly (a) does not need any proof. In order to prove (b), it is enough to prove $(\odot)$. First recall observation (iii) made during the earlier discussion of counting in the Hindu-Arabic numeral system, namely, for any whole number n,

$$\textit{the sum of ten } \; 1\underbrace{00\ldots0}_{n}\textit{'s} \quad \textit{is} \quad 1\underbrace{00\ldots0}_{n+1}.$$

This can be expressed in the present notation as:

$$\begin{aligned} 10 \times 10^n &= \overbrace{1\underbrace{00\ldots0}_{n} + 1\underbrace{00\ldots0}_{n} + \cdots + 1\underbrace{00\ldots0}_{n}}^{10} \qquad (\text{by definition of } \times) \\ &= 1\underbrace{00\ldots0}_{n+1} \\ &= 10^{n+1} \end{aligned}$$

This proves one half of $(\odot)$, and the other half is easier:

$$\begin{aligned} 10[n+1] &= \overbrace{10 \times 10 \times \cdots \times 10}^{n+1} \\ &= 10 \times \overbrace{10 \times 10 \times \cdots \times 10}^{n} \\ &= 10 \times 10[n] \qquad (\text{by definition of } 10[n]) \end{aligned}$$

The proof of (b) (and therefore the proof of (3)) is complete.

We now have another expression of 10^n given by (3), namely, $10^n =$

$10 \times 10 \times \cdots \times 10$ (n times). This has obvious implications, e.g.,

$$
\begin{aligned}
10^1 \times 10^3 &= 10^4 \\
10^2 \times 10^4 &= 10^6 \\
10^3 \times 10^4 &= 10^7 \\
10^4 \times 10^5 &= 10^9.
\end{aligned}
$$

This suggests that in general, if m and n are any whole numbers,

$$10^m \times 10^n = 10^{m+n} \tag{4}$$

Because (3) is available to us, we can dispatch (4) easily enough. If $m = 0$ or $n = 0$, there is nothing to prove, so we may assume $m, n > 0$. Then:

$$
\begin{aligned}
10^m \times 10^n &= \underbrace{10 \times 10 \times \cdots \times 10}_{m} \times \underbrace{10 \times 10 \times \cdots \times 10}_{n} \\
&= \underbrace{10 \times 10 \times \cdots \times 10}_{m+n} \\
&= 10^{m+n}
\end{aligned}
$$

So we see that (4) is true.[4]

It is a point well worth making that, without (3), it is not at all clear why (4) should be true. Remember, 10^m is *by definition* the number $100\ldots00$ (m zeros), so the left side of (4) means (by the definition of multiplication in (2)) adding $100\ldots0$ (n zeros) to itself $100\ldots00$ (m zeros) times. If n is a big number, such as 21334658, and $m = 68739$, can you be absolutely certain that adding $100\ldots0$ (21334658 zeros) to itself $100\ldots00$ (68739 zeros) times would result in the number

$$1000\ldots00 \ ((68739 + 21334658) \text{ zeros}) \ ?$$

So the key point of the proof of (4) is that, by using (3), the evaluation of the product $10^m \times 10^n$ is reduced to the counting of the number of copies of 10's, which we could do without any difficulty.

[4] But see the discussion of the associative law of multiplication in §2.

We are now in a position to explain the fact that multiplying 2,133,070 by 100,000 gives a number which can be obtained from 2,133,070 by adding the 5 zeros of 100,000 to the right of 2,133,070. That is,

$$2,133,070 \times 100,000 = 213,307,000,000$$

Note that, appearance notwithstanding, this fact is again *not* obvious! According to (2), the left side means adding 100,000 to itself 2,133,070 times. The last 5 digits of the resulting number will be 0's, to be sure, but why are the beginning digits exactly 2133070 ? The reason is:

$$\begin{aligned}
2,133,070 \times 100,000 &= \{ (2 \times 10^6) + (1 \times 10^5) + (3 \times 10^4) \\
&\quad +(3 \times 10^3) + (7 \times 10^1) \} \times 10^5 \\
&= (2 \times 10^6 \times 10^5) + (1 \times 10^5 \times 10^5) + \\
&\quad (3 \times 10^4 \times 10^5) + (3 \times 10^3 \times 10^5) + \\
&\quad (7 \times 10^1 \times 10^5) \\
&= (2 \times 10^{11}) + (1 \times 10^{10}) + (3 \times 10^9) + \\
&\quad (3 \times 10^8) + (7 \times 10^6) \qquad \text{(using (4))} \\
&= 213,307,000,000
\end{aligned}$$

The same reasoning of course allows us to write 213,307,000,000 more simply as $213,307 \times 10^6$. This is, incidentally, our first substantive application of the expanded form of a number,[5] but it will hardly be the last.

In a similar way, we can now write the number of seconds in a year as $31,536 \times 10^3$, and the age of the universe as 14×10^9 years. Furthermore, astronomers use as their unit of measurement a light year, which is 5,865,696,000,000 miles. We can now write this number as $5,865,696 \times 10^6$ miles.

The above reasoning is sufficiently general to show that if N is any nonzero whole number and k is any whole number, then

> $N \times 10^k =$ *the whole number obtained from N by attaching k zeros to the right of the last digit of N.*

[5] The discerning reader would have noticed that the above derivation made implicit use of the distributive law, which will be discussed presently in the next section.

Finally, we deal with the issue of why the number 830159 is the same number as 0830159, and is the same number as 000830159, etc. This is very easy to explain because by the expanded form of a number,

$$\begin{aligned} 000830159 &= 0 \times 10^8 + 0 \times 10^7 + 0 \times 10^6 + 8 \times 10^5 + \\ &\quad 3 \times 10^4 + 1 \times 10^2 + 5 \times 10^1 + 9 \times 10^0 \\ &= 830159\,. \end{aligned}$$

This observation is conceptually important in the understanding of the various algorithms in §3.

Exercise 1.1 Imagine you have to explain to a fourth grader that $43 \times 100 = 4300$. How would you do it?

Exercise 1.2 Imagine you have to explain to a fifth grader why $48 \times 500,000,000 = 24,000,000,000$. How would you do it?

Exercise 1.3 What number should be added to 946,722 to get 986,722? What number should be added to 68,214,953 to get 88,214,953?

Exercise 1.4 What number should be added to 58×10^4 to get 63×10^4? What number should be taken from 52×10^5 to get 48×10^5 ?

Exercise 1.5 Which is bigger? 4873 or 12001? 4×10^5 or 3×10^6? 8×10^{32} or 2×10^{33}? 4289×10^7 or 10^{11}? 765,019,833 or 764,927,919? Explain.

Exercise 1.6 Write each of the numbers 6100925730, 2300000000, and 7420000659 in expanded form.

Exercise 1.7 Show that for any nonzero whole number k, $10^k > m \times 10^{k-1}$ for any *single-digit* number m.

Exercise 1.8 The following is the introduction to the concept of *multiplication* taken from a third grade textbook. On the side of the page is the Vocabulary of the Day:

> MULTIPLICATION an operation using at least two numbers to find another number, called a product.
>
> PRODUCT the answer in multiplication.

Then in the text proper, one finds:

> How many are in 4 groups of 6?
>
> You can use MULTIPLICATION to solve the problem.
>
> Use cubes to model the problem and record the answer to the problem:

Number of groups		Number in each group		Product
6	$\times$	4	$=$	24

Further down:

> If Helena practices singing 3 hours each day for a week, how many hours will she practice altogether?
>
> Find: 3×7
> THERE IS MORE THAN ONE WAY!
>
> Method 1: You can use repeated edition to solve the problem.
>
> $$3 + 3 + 3 + 3 + 3 + 3 + 3 = 21$$
>
> Method 2: When the groups are equal, you can also write a MULTIPLICATION SENTENCE.
>
> $$7 \times 3 = 21.$$

Write down your reaction to the appropriateness of such an introduction, and compare your view with those of others' in your class.

2 The Basic Laws of Operations

This section discusses the basic laws which govern the *arithmetic* operations on whole numbers, e.g., why $23 + 79 = 79 + 23$, or why $(47 \times 4) \times 5 = 47 \times (4 \times 5)$. The key point of such a discussion is always that two collections of numbers which look superficially different are in fact equal, and this equality is indicated by the ubiquitous *equal sign* "=". Because the equal sign is one of the sources of confusion in elementary school, let us first deal with this symbol.

The most important thing to remember is that while the meaning of the equal sign does get more sophisticated as the mathematics gets more advanced, there is no reason for us to learn *everything* about this symbol all at once! We are starting from ground zero, the whole numbers, so the meaning of "=" is both simple and unequivocal: two numbers a and b are said to be *equal* if we can verify *by counting* that they are the same number.

For example, $4 + 5 = 2 + 7$ because we count 4 objects and then 5 more and get 9, whereas we count 2 objects and then 7 more and also get 9, so this is what $4 + 5 = 2 + 7$ means. So be sure to explain to your students — again and again if necessary — that the equal sign between two collection of whole numbers does *not* signify "do an operation to get an answer". It merely means:

> *check the numbers on both sides of the equal sign* by counting *to verify that both sides yield the same number.*

When we deal with fractions (Chapter 2) and decimals (Chapter 4), then obviously we can no longer just *count*. The equality of two fractions or two decimals will have to be more carefully explained (see §1 of Chapter 2 and §§1 and 3 of Chapter 4).

Now we come to the main concerns of this section: the associative laws and commutative laws of addition and multiplication, and the distributive law. These are without doubt among the most hackneyed items you have ever come across in mathematics. Your textbooks mention them with a sense of *noblesse obligé* and can't wait to get it over with, presumably they believe you deserve better. Yet we are going to spend the next twenty pages discussing exactly these laws without any apology. You are entitled to know why. There are at least three reasons. The first is that they are used everywhere, sometimes implicitly without your being aware of them. To bring home this point, let me just cite two of many instances where we have used them in §1 in an "underhanded" manner: the proof that $2,133,070 \times 100,000 = 213,307,000,000$ implicitly made use of the distributive law, and the proof of (4) (that $10^m \times 10^n = 10^{m+n}$) made implicit use of the associative law of multiplication, as we shall explain below (see the discussion above (10)). Because these laws infiltrate every aspect of arithemtic operations, your awareness of their presence would help you avoid making incorrect pronouncements. Take, for instance, the simple problem of addition: $12 + 25 + 18$. A quick way to compute this sum is to add 12 and 18 to get 30, and then add 30 to 25 to get 55. Now, some teachers probably explain this shortcut to their second graders by saying

> you just "fly over" the middle number and add the two at both ends,

but this would be incorrect because **by convention**,

> *in any expression involving arithemtic operations such as* $12 + 25 + 18$ *or* $4 \times 17 \times 25$*, it is understood that one only adds or multiplies two* neighboring *numbers at a time: e.g.,* $12 + 25$*, or* $25 + 18$*, or* 4×17*, or* 17×25*.*

To add or multiply two number by "flying over" another number is not a permissible move in mathematics. But the possibility of computing $12+25+18$ as $(12+18)+25 = 30+25 = 55$ or $4\times17\times25$ as $17\times(4\times25) = 17\times100 = 1700$ will turn out to be entirely correct, because the apparent "flying over" can be precisely justified by the commutative and associative laws of addition, and the commutative and associative laws of multiplication, respectively. As a teacher, you have to be ready with the correct mathematical explanations for such phenomena when the occasion calls for it.

A second reason is that knowing these basic laws *can* be helpful. Consider the following simple problems:

(a) $(87169 \times 5) \times 2 = ?$
(b) $10^7 \times 6572 = ?$
(c) $4 \times ([25 \times 18] + [7 \times 125]) = ?$

These computations can be tedious if taken literally. For example, the definition in (2) of §1 implies that (b) must be computed by adding 6572 to itself $10,000,000$ times. Yet with a judicious application of the basic laws, these computations can all be done effortlessly in one's head so that the answers are, respectively, 871690, 65,720,000,000, and 5300. See the later discussions in this section for the explanations.

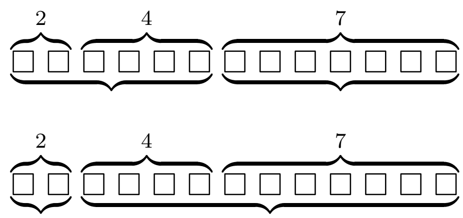
A final reason, and the most substantive one for our purpose, is that these laws play a central role in this monograph. They weave in and out of the five chapters, but are especially prominent in §§3.3–3.4 of Chapter 1, §7.2 of Chapter 2, and §3 of Chapter 5. For this reason, we take the opportunity to clarify these foundational matters once and for all.

For addition, the two basic laws are the *associative law* and the *commutative law*. These state that, if L, M and N are whole numbers,

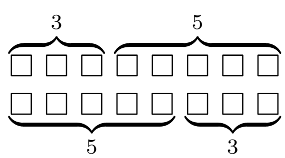
$$(L + M) + N = L + (M + N) \qquad (5)$$

$$M + N = N + M \qquad (6)$$

These are nothing more than summaries of our collective experiences with whole numbers, and **we take their truth as an article of faith**.[6] For illustrations of some of these experiences, think of the addition of two numbers as combining two groups of discrete objects. Then the following pictures describe the validity of the associative law (5) for $L = 2$, $M = 4$ and $N = 7$:



It is obvious from the pictures that whether we combine 2 and 4 first and then combine the sum with 7, or combine 2 with the sum of 4 and 7, we get 13. Needless to say, the pictorial evidence for the truth of (5) can be similarly obtained for other numbers. As to the commutative law (6), the following pictures show that whether we combine 3 objects with 5 objects or 5 objects with 3 objects, either way we get 8 objects:



Again, the pictorial evidence is not dependent on the particular choice of $M = 3$ and $N = 5$ and would hold equally well for other numbers.

We pause to comment on the equal sign in the context of the associative law (5) to reinforce the definition of this symbol. What (5) asserts is that the two numbers $(L + M) + N$ and $L + (M + N)$ are the same. In other words, we count the left side by first counting the group consisting of L objects and M objects, and then continue the counting to include the next N objects; we get one number. For the right side, we begin with L objects, then continue the counting to include the group consisting of M objects and N objects, thereby getting a second number. What (5) says is that these two numbers are the same number.

[6] In the proper axiomatic setup, both of (5) and (6) can be proved as theorems.

One consequence of the associative law of addition is that it clears up the meaning of such common expressions as $4 + 3 + 7$. This triple addition is a priori ambiguous, because it could mean either $(4 + 3) + 7$ or $4 + (3 + 7)$. (Note again that adding 4 to 7 *before* adding the result to 3 is not an option because, by convention, only neighboring numbers are added.) But (5) tells us that there is in fact no ambiguity because the two ways of adding are the same. In general then, (5) leads to the conclusion that we don't need to use parentheses in writing the sum of any three whole numbers $l + m + n$, e.g., $4 + 3 + 7$. Once this is noted, however, it should not be surprising that we can draw the same conclusion about the sum of any four whole numbers $l + m + n + p$. For example, let us show:

$$((l + m) + n) + p = l + ((m + n) + p),$$

where

> *the* **convention** *regarding parentheses is to do the innermost parentheses first and then systematically work one's way out. Thus* $((l + m) + n) + p$ *means: add* l *to* m*, then add the result to* n*, and finally add* $((l + m) + n)$ *to* p*.*

Let us explain the preceding equality of four numbers by using (5) repeatedly. Letting $L = (l + m)$, $M = n$, and $N = p$ in (5), we obtain

$$((l + m) + n) + p = (l + m) + (n + p). \qquad (\natural)$$

Now let $L = l$, $M = m$, and $N = (n + p)$ in (5) again, then the right side of $(\natural)$ becomes

$$(l + m) + (n + p) = l + (m + (n + p)).$$

Finally, letting $L = m$, $M = n$, and $N = p$ and reading (5) from right to left, we obtain $m + (n + p) = (m + n) + p$. Substituting this value of $m + (n + p)$ into the right side of the preceding equation, we obtain:

$$(l + m) + (n + p)) = l + ((m + n) + p). \qquad (\sharp)$$

Putting $(\natural)$ and $(\sharp)$ together, we get

$$((l + m) + n) + p = l + ((m + n) + p),$$

as desired. Consequently, the meaning of $l + m + n + p$ is also unambiguous without the use of parentheses.

Because this kind of abstract, formal reasoning looks so facile and believable, there is the danger that you would take it for granted. Let me therefore make sure you are aware that there is substance beneath the smooth surface of formalism. For example, if we let $l = 5$, $m = 2$, $n = 8$ and $p = 1$ and compute both $((5+2)+8)+1$ and $5+((2+8)+1)$ directly to compare the results step-by-step, it is far from obvious why they turn out to be equal at the end:

$$\begin{array}{rclclcl} ((5+2)+8)+1 &=& (7+8)+1 &=& 15+1 &=& 16 \\ 5+((2+8)+1) &=& 5+(10+11) &=& 5+11 &=& 16 \end{array}$$

This is the first instance that we come across the paradoxical situation that, by resorting to formal abstract reasoning, an argument for the general case (which should be more difficult) turns out to be more understandable than the special case. I want to assure you that this will not be the last instance where this happens.

The same reasoning shows that

given any collection of numbers, say 26 of them $\{a,\ b,\ c, \cdots, y,\ z\}$, *we can unambiguously write*

$$a + b + c + \cdots + y + z$$

without the use of any parentheses and, regardless of the order of the addition of the numbers,[7] *the result will be the same.*

You cannot fail to notice at this point that in the very first sum of §1, i.e., $20000 + 500 + 40 + 1$, we have already made implicit use of this fact. The same can be said of the later expression (in §1) that

$$31536000 = 30000000 + 1000000 + 500000 + 30000 + 6000.$$

Needless to say, there are many other examples of using the associative law of addition without mentioning it.

Essentially the same comments apply to the commutative law (6). For example, if we have three numbers l, m and n, then all six expressions

$$\begin{array}{lll} l+m+n & l+n+m & m+l+n \\ m+n+l & n+l+m & n+m+l \end{array}$$

[7] Don't forget the convention that one can only add two neighboring numbers at a time.

are the same. (Notice that we have already made use of the fact that the use of parentheses is not needed for the addition of three numbers !) Let us show, for instance, that $l+m+n = n+m+l$ by applying (6) repeatedly to two numbers at a time:

$$l+m+n \;=\; m+l+n \;=\; m+n+l \;=\; n+m+l$$

Of course the same is true of the addition of any collection of whole numbers:

> *the order of appearance of the whole numbers in any finite sum is unimportant.*

By now, you undoubtedly appreciate one aspect of the laws (5) and (6): they are tremendous labor saving devices. If we have five numbers 2, 4, 5 11, 3, then there are 120 ways of adding them (try it!). But we only have to do one of them, say, $((4+5)+11)+(2+3) = (9+1)+5 = 20+5 = 25$, and there is no need to do any of the other 119 sums because they would all be equal to 25. This then justifies the writing of $2+5+4+11+3 = 25$.

Before proceeding further, it may be time to tie up the loose end left open earlier as to why it is permissible to first add $12+18$ in the sum $12+25+18$. In detail, we are arguing that

$$12+25+18 = (12+18)+25.$$

Observe first of all that the associative law of addition is already used in writing $12+25+18$ without the use of parentheses; we may insert parentheses any which way we wish. So with this understood, we give the proof:

$$\begin{aligned} 12+25+18 &= 12+(25+18) & & \\ &= 12+(18+25) & & \text{(commutative law)} \\ &= (12+18)+25 & & \text{(associative law)} \end{aligned}$$

Exactly as claimed.

We should emphasize that argument of this sort is primarily for your benefit as a teacher and should not be taken as a statement that good teaching means *always* explaining details of this kind. What it does say is that, as a teacher, you should be ready with an explanation. In the fourth or fifth grade, for instance, there will be occasions for you to introduce this kind of reasoning to your students, gently. But a teacher must exercise good judgment in not overdoing anything.

Next we turn to the multiplication of whole numbers as defined in (2) of §1. We have two similar laws, the *associative law of multiplication* and the *commutative law of multiplication*: for for any whole numbers L, M and N,

$$(LM)N = L(MN) \tag{7}$$
$$MN = NM \tag{8}$$

where we have made use of the

> **notational convention:** *when letters are used to stand for numbers, the multiplication sign is omitted so that "MN" stands for "$M \times N$".*

Before discussing the empirical evidence behind (7) and (8), we can demonstrate their power by applying them to problems (a) and (b) posed at the beginning of this section. First, $(87169 \times 5) \times 2$ is equal to $87169 \times (5 \times 2)$, by the associative law (7). Since 5×2 is 10, the triple product is immediately seen to be equal to 871690. This disposes of (a). Next, $10^7 \times 6572 = 6572 \times 10^7$ by the commutative law (8), but the right side is 65,720,000,000 by the conclusion we arrived at near the end of §1, to the effect that

> $N \times 10^k =$ *the whole number obtained from N by attaching k zeros to the right of the last digit of N.*

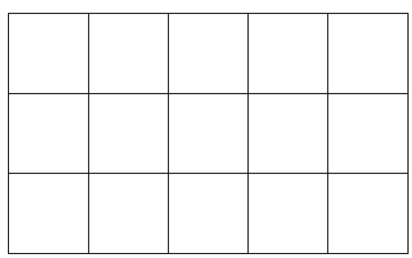
This then disposes of (b).

The way we have just dealt with (a) and (b) affords an excellent opportunity to remind you of the importance of using definitions *exactly as given*. It would have been very easy, for example, for you to equate in your minds *without thinking* that $10^7 \times 6572 = 6572 \times 10^7$ and then conclude that the result is 65,720,000,000. What we are doing here is to intentionally bring the underlying reason (the commutative law (8)) for the validity of $10^7 \times 6572 = 6572 \times 10^7$ to your attention: by definition, $10^7 \times 6572$ is adding 6572 to itself 10,000,000 times, whereas 6572×10^7 is adding 10,000,000 to itself 6572 times. They look at first glance, at least, to be quite different animals. The equality of these two products may not seem important to you because you have taken it for granted. But if you do not begin to take note of the role of (8) in arithmetic, you would not be able to follow the reasoning in later discussions when the direction of the discussion is determined by (8), e.g., in §3.4 when we show that the two interpretations of division yield the same result.

As with the corresponding laws for addition, one may regard (7) and (8) as summaries of empirical experiences and accept them on faith. Here is the kind of pictorial evidence that is easily available:

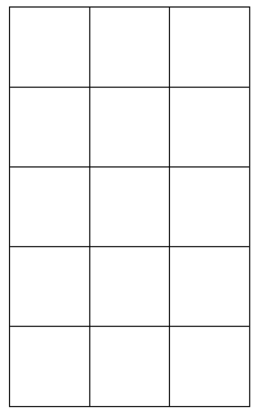
> **Activity:** Verify by direct counting that (8) is valid for M and N between 1 and 5 inclusive (in symbols: $1 \leq M, N \leq 5$), by using rectangular arrays of dots to represent multiplication of whole numbers as in the discussion below (2) in §1.
>
> **Activity:** Verify by direct counting that (7) is valid for L, M, N between 1 and 5 inclusive (in symbols: $1 \leq L, M, N \leq 5$), by using 3-dimensional rectangular arrays of dots to represent triple products of whole numbers.

For later needs in this monograph (cf. the discussions of fractions and decimals in Chapters 2 and 4), the representation of multiplication by dots would be inadequate and it would be more suitable to use a *area model*. For example, 3×5 is represented as a rectangle with "vertical" length 3 (corresponding to the first number) and "horizontal" length 5 (corresponding to the second number):



(9)

According to this description of the area model, the product 5×3 would be represented by the area of a rectangle with "vertical" length 5 and "horizontal" length 3:



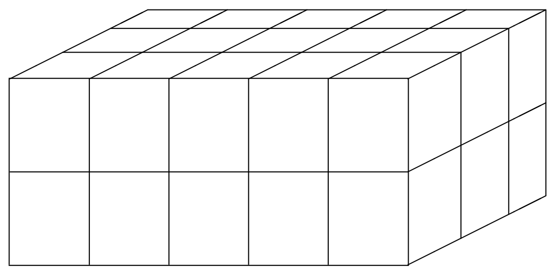
Because these two rectangles can be obtained from each other by a 90° rotation, they have the same area. This then give a pictorial “explanation” of $3 \times 5 = 5 \times 3$. As usual, this discussion is independent of the choice of the numbers 3 and 5 and would be the same for any two numbers m and n.

> In §4, we will present yet another interpretation of the multiplication of whole numbers that will be important when we come to fractions. We cannot give this interpretation here because it involves a deeper understanding of what a “number” is.

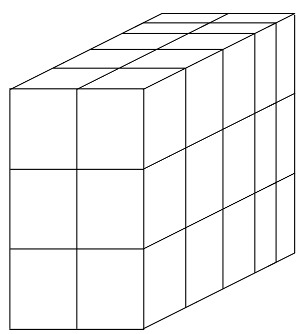
A word about “area” would be appropriate: A detailed discussion of this concept will not be attempted here, nor will it be needed, except to mention that, by convention, we agree to let the area of the *unit square* (the square with each side equal to 1) to be just 1, so that the area of the rectangle in (9) is 15 because precisely 15 unit squares tile (or pave) the rectangle.

It is worthwhile to point out that the area model of multiplication provides the mathematical underpinning of the manipulative *Base Ten Blocks*. Most likely you have used this manipulative in your classroom to facilitate the learning of multiplication. We would like to add a passing comment that while Base Ten Blocks, like other manipulatives, can be helpful, students should not be allowed to become dependent on it. The real challenge in a mathematics classroom is still to learn the mathematics, not manipulate manipulatives.

The product of three whole numbers l, m and n can be represented as the volume of a rectangular solid. For example, $(3 \times 5) \times 2$ is the volume of the rectangular solid of height 2 built on the rectangle in (9) above:



Similarly, $3 \times (5 \times 2) = (5 \times 2) \times 3$ by the commmutative law (8), so that $3 \times (5 \times 2)$ is the volume of the solid of height 3 built on the rectangle with "verticle" length 5 and "horizontal" length 2:



The equality of the volumes of these two solids — because one is obtained from the other by a rotation in space — is then the pictorial evidence for the truth of $(3 \times 5) \times 2 = 3 \times (5 \times 2)$. We remark as in the case of area that we shall not go into the precise definition of volume but will only use it in an intuitive way. Again, the concrete numbers 3, 5, and 2 of the preceding argument can be replaced by any triple of numbers in (7).

Of course, as in the case of addition, there are more general versions of (7) and (8) for arbitrary collections of numbers. For example, the previous discussion of the associative law and commutative law for addition ((5) and (6)) applies to multiplication verbatim, and we know that

> *the product of any collection of numbers can be written unambiguously without the use of parentheses and without regard to order.*

Thus, for any four numbers, say l, m, n, p, their product can be written in any of the following 24 ways:

$$\begin{array}{cccccc} lmnp & lmpn & lnmp & lnpm & lpmn & lpnm \\ mlnp & mlpn & mnlp & mnpl & mpln & mpnl \\ nlmp & nlpm & nmlp & nmpl & nplm & npml \\ plmn & plnm & pmln & pmnl & pnlm & pnml \end{array}$$

and all of them are equal to $((lm)n)p$. To drive home the point that first surfaced in connection with the associative law of addition, let us use four explicit numbers — say $l = 7$, $m = 3$, $n = 2$ and $p = 4$ — to illustrate the nontrivial nature of, for example, $mlpn = plnm$:

$$\begin{array}{rcrcrcr} (7 \times 3) \times (2 \times 4) & = & 21 \times (2 \times 4) & = & 21 \times 8 & = & 168 \\ (4 \times (7 \times 2)) \times 3 & = & (4 \times 14) \times 3 & = & 56 \times 3 & = & 168\,. \end{array}$$

At the risk of harping on the obvious, note that none of the intermediate steps of the two computations look remotely similar, and yet miraculously the final results are identical.

Now we are in a position to point out in what way the proof of (4) in §1, namely, $10^m \times 10^n = 10^{m+n}$, implicitly made use of the associative law of multiplication. Properly speaking, its proof should go as follows: Because we may write a product of m numbers without the use of parentheses, we have $10^m = \underbrace{10 \times 10 \times \cdots \times 10}_{m}$. Similarly, $10^n = \underbrace{10 \times 10 \times \cdots \times 10}_{n}$. Therefore,

$$10^m \times 10^n = (\underbrace{10 \times 10 \times \cdots \times 10}_{m}) \times (\underbrace{10 \times 10 \times \cdots \times 10}_{n})$$

However, we also know that a product of $m + n$ numbers can be written without the use of parentheses. Consequently,

$$10^m \times 10^n = \underbrace{10 \times 10 \times \cdots \times 10}_{m+n},$$

and the right side is just 10^{m+n}. This shows $10^m \times 10^n = 10^{m+n}$.

Finally, the *distributive law* connects addition with multiplication. It states that for any whole numbers M, N, and L,

$$M(N+L) \;=\; MN+ML \tag{10}$$

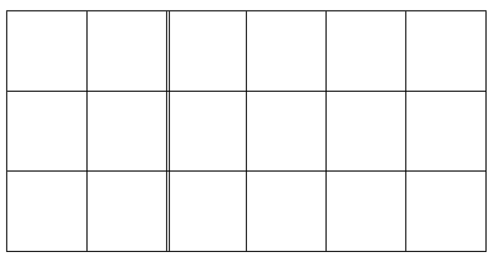
Recall in this connection

> *the* **convention** *that in the expression $MN+ML$, we multiply the numbers MN and ML first before adding.*

Again, like the other laws we have discussed so far, the distributive law (10) is nothing more than a summary of common sense. Pictorially, we have:

> **Activity:** Directly verify (10) using rectangular arrays of dots to represent multiplication for $1 \le L, M, N \le 5$.

One can also use the area model: if $M=3$, $N=2$ and $L=4$, then $3(2+4)$ is the area of the following big rectangle:



On the other hand, 3×2 is the area of the left rectangle and 3×4 is the area of the right rectangle. Thus $3(2+4) = (3\times 2)+(3\times 4)$. Again, the essence of this picture is unchanged when 2, 3, and 4 are replaced by other triples of numbers.

Clearly, *the distributive law generalizes to more than three numbers.* For example:

$$m(a+b+c+d) \;=\; ma+mb+mc+md$$

for any whole numbers m, a, b, c and d. This can be seen by applying the distributive law (10) twice, as follows:

$$\begin{aligned} m(a+b+c+d) &= m(\{a+b\}+\{c+d\}) \\ &= m\{a+b\}+m\{c+d\} \\ &= \{ma+mb\}+\{mc+md\} \\ &= ma+mb+mc+md \end{aligned}$$

Observe that each step in the preceding calculation depends on the earlier discussion of the legitimacy of writing $a+b+c+d$ without parentheses.

As mentioned earlier, the distributive law is the glue that binds addition $+$ and multiplication $\times$ together. Despite its obvious importance, it seems to be the least understood among the three laws and for many, a firm grasp of this law proves elusive. We urge you to spare no effort in learning to use it effectively. Spending more time with some of the exercises at the end of this section may be the answer.

A common mistake in connection with the distributive law is to remember (10) only in the form of "the left side of (10) is equal to the right side" (i.e., $M(N+L) = MN + ML$) without realizing that (10) also says "the right side of (10) is equal to the left side" (i.e., $MN + ML = M(N+L)$). In other words, given $35 \times (72+29)$, most people recognize that it is equal to $(35 \times 72) + (35 \times 29)$, but when $(35 \times 72) + (35 \times 29)$ is given instead, they fail to see that it is equal to $35 \times (72+29)$. In practice, the latter skill may be the more critical of the two, and there is an mathematical reason for this. In terms of the preceding example, $(35 \times 72) + (29 \times 35)$ involves *two* multiplications (35×72 and 29×35) and one addition, whereas $35 \times (72+29)$ involves only *one* multiplication and one addition. Because multiplication is in general more complicated, it is preferable to multiply as little as possible and therefore preferable to compute $35 \times (72+29)$ rather than $(35 \times 72) + (29 \times 35)$. Therefore, to recognize $(35 \times 72) + (29 \times 35) = 35 \times (72+29)$ is to be able to achieve a simplification.

To further pin down the last idea, we bring closure to this section by doing problem (c) posed near the beginning: $4 \times (25 \times 18 + 7 \times 125) = ?$ We know $125 = 5 \times 25$, so that $7 \times 125 = 7 \times 5 \times 25 = 25 \times [7 \times 5]$. Therefore:

$$\begin{aligned} 4 \times \big(25 \times 18 + 7 \times 125) &= 4 \times (25 \times 18 + 25 \times [7 \times 5]\big) \\ &= 4 \times 25 \times (18 + [7 \times 5]) \qquad \text{(distributive law)} \\ &= 100 \times 53 = 53 \times 100 \\ &= 5300. \end{aligned}$$

It should be clear that the whole computation can be done by mental math. On the other hand, we would be looking at calculating three multiplications if we don't appeal to the distributive law: 25×18, 7×125, and 4×1325, where $1325 = 25 \times 18 + 7 \times 125$.

Moral: Be sure you know $MN + ML = M(N + L)$.

We began the discussion of *order* among whole numbers (i.e., which whole number is bigger) in §1. We can now conclude that discussion. Recall that given two whole numbers a and b, the inequality $a < b$ is defined to mean that in the counting of the whole numbers starting with 0, 1, 2, ..., the number a precedes b. For the convenience in logical arguments, we wish to express this definition differently:

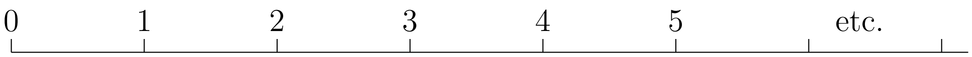
> *The statement $a < b$ is the same as the statement: there is a nonzero whole number c so that $a + c = b$.*

Before giving the reason for this assertion, we explain what is meant by the two conditions being *the same*. This is a piece of mathematical terminology that signifies that both of the following statements are true:

> *If $a < b$, then there is a nonzero whole number c so that $a + c = b$.*
>
> *Conversely, if there is a nonzero whole number c so that $a+c = b$, then $a < b$.*

In concrete situations, both statments are quite obvious. For example, $7 < 12$ means that in counting from 7, we have to go 5 more steps before we get to 12, so $7 + 5 = 12$. Conversely, if we know $7 + 5 = 12$, then we must go 5 more steps from 7 before we get to 12, so $7 < 12$. The general reasoning is not much different. Given $a < b$, we know by definition that a precedes b, so that in the counting of the whole numbers, after we get to a, we need to go (let us say) c more steps before getting to b, and c is not zero. This implies $a + c = b$. Conversely, suppose $a + c = b$ is given, with $c > 0$, then after counting a objects, we have to count c more objects before we get b objects. So by the definition of "smaller than", we know $a < b$.

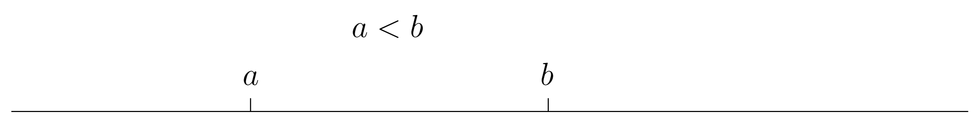
At this juncture, it is time to introduce the *number line* in order to give a geometric interpretation of inequalities. Fix a straight line and designate a point on it as 0. To the right of 0, mark off equally spaced points and call them 1, 2, 3, 4, etc., as on a ruler. Thus the whole numbers are now identified with these equally spaced points on the right side of the line. It is convenient to single these points out by *markers* (notches), again as on a ruler, and to identify the points with the markers. We explicitly call attention to the fact that the counting of the whole numbers (as was done in §1) corresponds to the progression of the markers to the right of the line starting with the initial marker 0, as shown below. (Until Chapter 5, we will have no need for the part of the line to the left of 0):



It follows from the way the number line is drawn that

> *two whole number a and b satisfy $a < b$ precisely when the position of a on the number line is to the left of that of b.*

Here is a pictorial representation of the situation:



A further geometric interpretation can be given: given $a < b$ as shown, suppose $a + c = b$, then c *is precisely the number of markers between a and b.* This follows immediately from the way the whole numbers are positioned on the number line.

> **Activity:** Verify the last statement about c for some concrete numbers such as $a = 8$, $b = 21$, or $a = 86$, $b = 95$.

The following facts about inequalities are well-known:

$$\begin{array}{rcl} a+b < a+c & \text{is the same as} & b < c \\ a \neq 0,\ ab < ac & \text{is the same as} & b < c \\ a < b \text{ and } c < d & \text{implies} & a+c < b+d \\ a < b \text{ and } c < d & \text{implies} & ac < bd \end{array} \qquad (11)$$

We will only give the explanation for the second assertion (the most difficult of the four) and leave the rest as exercises.

> **Activity:** Convince yourself that $2a < 2b$ is the same as $a < b$, and that $3a < 3b$ is the same as $a < b$. Do it both numerically as well as on the number line.

Now the proof of the second assertion. Given $a \neq 0$. First we prove that

$$ab < ac \text{ implies } b < c.$$

There are three possibilities between the whole numbers b and c: (A) $b = c$, (B) $c < b$, and (C) $b < c$. We know ahead of time that only one of the three possibilities holds. In order to show that (C) is the correct conclusion, all we need to do is to show that both (A) and (B) are impossible. Now if (A) holds, then $ab = ac$, which is contrary to the assumption that $ab < ac$. We have therefore eliminated (A). If (B) holds, then $c < b$, and there will be a nonzero whole number l so that $c + l = b$. It follows that $a(c + l) = ab$, which implies $ac + al = ab$, by the distributive law. But both a and l are nonzero, so $al > 0$ and therefore $ac < ab$. This is also contrary to the assumption that $ab < ac$. Hence (B) is also eliminated. It follows that (C) is the only possible conclusion.

Next, we prove the converse, i.e.,

$$b < c \text{ implies } ab < ac.$$

Now $b < c$ implies that $b + l = c$ for some nonzero l. Thus $a(b + l) = ac$, so that $ab + al = ac$ again by the distributive law. Because al is nonzero, the last equation means $ab < ac$, as desired. This completes the proof.

Exercise 2.1 Elaine has 11 jars in each of which she put 16 ping pong balls. One day she decided to redistribute all her ping pong balls equally among 16 jars instead. How many balls are in each jar? Explain.

Exercise 2.2 Before you get too comfortable with the idea that everything in this world has to be commutative, consider the following. (i) Let A_1 stand for "put socks on" and A_2 for "put shoes on", and let $A_1 \circ A_2$ be "do A_1 first, and then A_2", and similarly let $A_2 \circ A_1$ be "do A_2 first, and then A_1". Convince yourself that $A_1 \circ A_2$ does not have the same effect as $A_2 \circ A_1$. (ii) For any whole number k, let $B_1(k)$ be the number obtained by adding 2 to k, and $B_2(k)$ be the number obtained by multiplying k by 5. Show that no matter what the number n may be, $B_1(B_2(n)) \neq B_2(B_1(n))$.

Exercise 2.3 Find shortcuts to do each of the following computations and give reasons (associative law of addition? commutative law of multiplication? etc.) for each step: (i) $833 + (5167 + 8499)$, (ii) $(54 + 69978) + 46$, (iii) $(25 \times 7687) \times 80$, (iv) $(58679 \times 762) + (58679 \times 238)$, (v) $(4 \times 4 \times 4 \times 4) \times (5 \times 5 \times 5 \times 5 \times 5)$, and (vi) 64×125, (vii) $(69 \times 127) + (873 \times 69)$, M (viii) $\big(([125 \times 24] \times 674) + ([24 \times 125] \times 326)\big)$.

> The purpose of the last exercise is not to get you obsessed with tricks in computations everywhere. Tricks are nice to have, but they are not the main goal of a mathematics education, contrary to what some people would have you believe. What this exercise tries to do is, rather, to make you realize that the basic laws of operation discussed in this section are more than empty, abstract gestures. They have practical applications too.

Exercise 2.4 Prove the remaining three assertions in (11).

Exercise 2.5 Prove that both assertions in (11) remain true if the strict inequality symbol "$<$" is replced by the weak inequality symbol "$\leq$".

Exercise 2.6 Let m and n be a 3 digit number and a 2 digit number, respectively. Can mn be a 4 digit number? 5 digit number? 6 digit number? 7 digit number?

Exercise 2.7 Let m and n be a k digit number and an ℓ digit number, respectively, where k and ℓ are nonzero whole numbers. How many digits can the number mn be? List all the possibilities and explain.

Exercise 2.8 Suppose you have a calculator which displays only 8 digits (and if you have a fancy calculator, you will be allowed to use only 8 digits!), but you have to calculate 856164298×65. Discuss an efficient method to make use of the calculator to help with the comutation. Explain. Do the same for 376241048×872.

Exercise 2.9 Let x and y be two whole numbers. (i) Explain why $(x+y)(x+y) = x(x+y) + y(x+y)$. (ii) Explain why $(x+y)(x+y) = xx + xy + yx + yy$. (iii) Explain why $(x+y)(x+y) = x^2 + 2xy + y^2$, where as usual x^2 means xx and y^2 means yy.

Exercise 2.10 Let x and y be two whole numbers. Explain why $(x-y)(x+y) = (x-y)x + (x-y)y$.

Exercise 2.11 The following is how a fourth grade textbook introduces the *associative law of multiplication*.

> Ramon buys yo-yos from two companies, He buys six different styles from each company and gets each style in 4 different colors. How many yo-yos does he buy in all?
>
> Find $2 \times 6 \times 4$ to solve. You can use the ASSOCIATIVE PROPERTY to multiply three factors. The grouping of the numbers does not affect the answer.
>
> Step 1: Use parentheses to show grouping.

$$2 \times 6 \times 4 = (2 \times 6) \times 4$$

Step 2: Look for a known fact to multiply.

2×4 is a known fact.

Step 3: Use the Commutative Property to change the order, if necessary.

$$\begin{aligned}(2 \times 6) \times 4 &= (6 \times 2) \times 4 \\ &= 6 \times (2 \times 4) \\ &= 6 \times 8 = 48\end{aligned}$$

Write down your reaction to such an introduction, and compare with those of others' in your class.

3 The Standard Algorithms

By an *algorithm*, we mean an explicitly defined, step-by-step computational procedure which has only a finite number of steps. The purpose of this section is to describe as well as provide a complete explanation of the so-called *standard algorithms* for the four arithmetic operations among whole numbers.

At the outset, we should make clear that there is no such thing as *the unique* standard algorithm for any of the four operations $+$, $-$, $\times$, $\div$, because minor variations in each step of these algorithms not only are possible, but have been incorporated into the algorithms by various countries and even different ethnic groups. The underlying mathematical ideas are however always the same, and it is these ideas that are the focus of our attention here. For this reason, the nomenclature of "standard algorithms" is eminently justified. This is not to say that the algorithms themselves — the mechanical procedures — are of no interest. On the contrary, they are, because computational techniques are an integral part of mathematics. Furthermore, the conciseness of these algorithms, especially the multiplication algorithm and the long division algorithm, is a marvel of human invention. One of the goals of this section is to make sure that you come away with a renewed respect for them. Nevertheless, we shall concentrate on the mathematical ideas behind them as you are likely to be less familiar with these ideas.

A fundamental question about these arithmetic algorithm is *why you should bother to learn them*. Take a simple example: what is 17×12 ? By definition, this is 12 added to itself 17 times and one school of thought would have you count 17 piles of birdseed with 12 in each pile. But what about $34,609 \times 549,728$? Because brutal counting is less than attractive for numbers of this magnitude, a shortcut is clearly called for. This is where algorithms come in: they provide a shortcut in lieu of direct counting. Therefore at the outset, the *efficiency* of an algorithm — how to get the answer as simply and quickly as possible — is an overriding concern. But one could push this argument further. Why worry about efficiency if pushing buttons on a calculator may be the most efficient way to make a computation such as $34,609 \times 549,728$? There are at least two reasons why a strict reliance on the calculator is inadequate. First, without a firm grasp of the place value of our numeral system and the mathematical underpinning of the algorithms, it would be impossible to detect mistakes caused by pushing the wrong buttons on the calculators.[8] A more important reason is that learning the reasoning behind these efficient algorithms is a very compelling way to acquire many of the fundamental skills in mathematics, including abstract reasoning and symbolic manipulative skills. These are skills absolutely essential for the understanding of fractions and decimals in the subsequent chapters. Much more is true. The present crisis in the learning of algebra in schools would have been largely eliminated if students were properly taught the logical reasoning behind the algorithms in the upper elementary and middle schools.[9] Please keep all this in mind when you learn the mathematics surrounding the algorithms and especially when you teach them to your students.

There is a kind of *leitmotif* all through the algorithms, which can be roughly described as follows:

> *To perform a computation with an n digit number, break the computation into n steps so that each step involves only one digit of the given number.*

The precise meaning of this statement will be made clear with each algorithm, but the overall idea is that those *simpler single-digit computations* can all be

[8] I trust that it would be unnecessary to recount the many horror stories related to finger-on-the-wrong-button.

[9] This conclusion is based on the fact that, in mathematics, learning does not take place without a solid foundation.

carried out *routinely without thinking*. The last sentence calls for some comments as it runs counter to some education theories which a segment of the education community holds dear. The fact that a crucial part of mathematics rests on the ability to break down whole concepts into discrete sub-concepts and sub-skills must be accepted if one hopes to learn mathematics. This is the very nature of mathematics, and no amount of philosophical discussion would change that. A second point concerns the uneasiness with which some educators eyes the *routine* and *nonthinking* nature of an algorithm. It is the very routineness that accounts for the fact that these algorithms get used; it guarantees an easy way to get results. If we teach these algorithms without emphasizing their routine character, we would be falsifying mathematics.

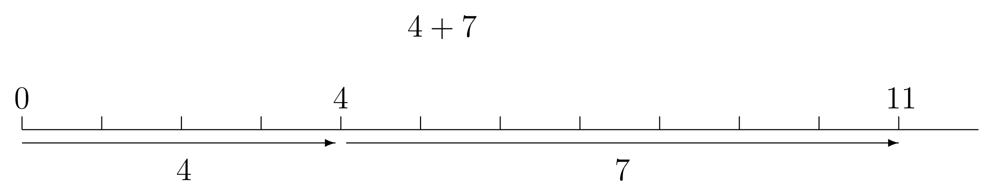
As to the *non-thinking* aspect of these algorithms, there is at present a perception that if anything can be done without thinking, then it does not belong in a mathematics classroom. This is WRONG. If mathematicians are forced to do mathematics by having to think *every* step of the way, then little mathematics of value would ever get done and all research mathematics departments would have to close shop. What is closer to the truth is that deep understanding of a topic tends to reduce many of its sophisticated processes to simple mechanical procedures. The ease of executing these mechanical procedures then frees up mental energy to make possible the conquest of new topics through imagination and mathematical reasoning. In turn, much of these new topics will (eventually) be once again reduced to routine or nearly routine procedures, and the process then repeats itself. There is nothing to fear about the ability to execute a correct mathematical procedure with ease, i.e., without thinking. More to the point, having such an ability in the most common mathematical situations is not only a virtue but an *absolute necessity*. What one must fear is limiting one's mastery of such procedures to only the mechanical aspect and ignoring the mathematical understanding of *why* the procedures are correct. A teacher's charge in the classroom is to promote both facility with procedures and the ability to reason. In the teaching of these algorithms, we should emphasize both their routine nature as well as the logical reasoning that lies behind the procedures.

The preceding discussion is about the kind of mathematical understanding teachers of mathematics must have in approaching the basic arithmetic algorithms. The pedagogical issue of how to introduce these algorithms to children in the early grades is something that lies outside the scope of this monograph and needs to be treated separately. There is however a discussion of this issue concerning the addition and multiplication algorithms in the article, "Basic skills versus conceptual understanding", in the Fall 1999 issue of the *American Educator*, p. 14; it is also accessible at

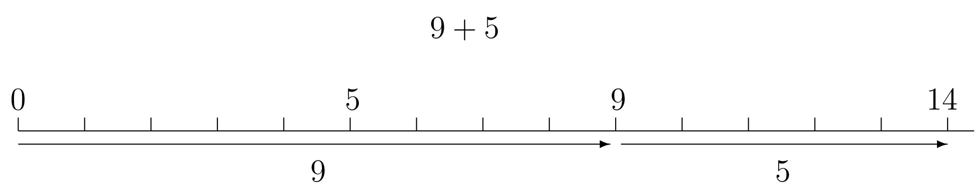
http://www.aft.org/american_educator/fall99/wu.pdf

3.1 An addition algorithm

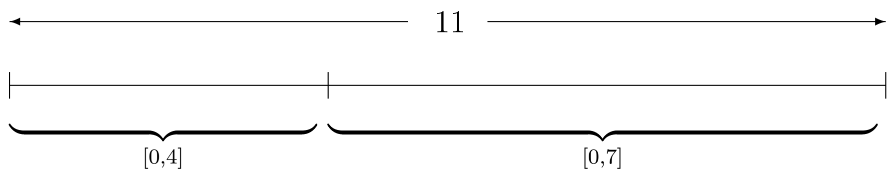
Given any two numbers, say 4 and 7, to find the sum $4+7$ is a simple matter in principle: start with 4, we count 7 more times until we reach 11, and that would be the sum. There is a graphic representation of this in terms of the number line:



Similarly, here is a graphic representation of $9+5=14$:

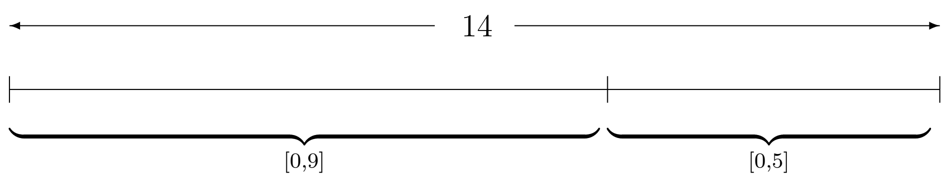


For many purposes, it is more convenient to take a slightly different point of view. Denote the line segment from 0 to 7 by $[0,7]$. It is natural to call 7 the *length* of $[0,7]$. Simlarly, we define the *length* of the segment $[0,n]$ from 0 to the whole number n to be n. We can now define the length of more general line segments by treating the number line as an "infinite ruler", as follows. For any line segment $[x,y]$ from a point x to another point y on the number line, — where it is understood from the notation $[x,y]$ that y is to the right of x — we say the *length* of $[x,y]$ is n for a whole number n if, by sliding $[x,y]$ to the left along the number line until x rests at 0, the right endpoint y rests at n. With the notion of length at our disposal, we can describe another way to find the sum of $4+7$: *concatenate* the two segments $[0,4]$ and $[0,7]$ in the sense of placing them on a straight line end-to-end, then the length of the concatenated segment is the sum of $4+7$. Thus:



Notice that concatenating the left endpoint of $[0,7]$ to the right endpoint of $[0,4]$ corresponds exactly to "start with 4, we count 7 more times until we reach 11".

Similarly, we can concatenate $[0,9]$ and $[0,5]$ to get $9+5=14$:



This discussion continues to be meaningful when $4+7$ and $9+5$ are replaced by $a+b$ for any $a+b$ where a and b are whole numbers. *In principle* then, addition is simple.

Now look at $4502+263$. It can be rather trying to count 263 times starting from 4502 before getting an answer. (Try it!) However, a special feature of the Hindu-Arabic numeral system, namely, the fact that its numerals "already come pre-packaged", renders such a desperate act completely unnecessary. Let us use a simpler example of addition to explain what is meant by the phrase in quotes. Suppose we have two sacks of potatoes, one containing 34 and the other 25, and we want to know how many potatoes there are altogether. One way is to dump the content of both sacks to the ground and start counting, which is what we called "the desperate act". But suppose upon opening the sacks, we find that in each sack, the potatoes come in bags of 10: the sack of 34 potatoes is put in 3 bags of 10 each plus 4 stray ones, while the sack of 25 is put in 2 bags of 10 each plus 5 stray ones. Therefore an intelligent way to count the total number of potatoes is to first count the total number of bags of 10's ($3 + 2$ is 5, so there are 5 bags of 10 each), and then count the stray ones separately ($4 + 5$ is 9, and so there are 9 extra). Thus the total is 5 bags of 10 each, plus 9 extra ones, which puts the total at 59. **This is exactly the idea behind the addition algorithm** because the number 3 in 34 — being in the tens place — signals that there are 3 tens, and the 2 in 25 signals that there are 2 tens. Adding 2 and 3 to get 5, we know that there are 5 tens in the total. Adding 4 to 5 then rounds off the whole sum, and we get $34 + 25 = 59$.

The standard *addition algorithm* is nothing more than a formal elaboration of this simple idea. It says:

> *The sum of two whole numbers can be computed by lining them up digit-by-digit, with their ones digits in the extreme right column, and adding the digits* column-wise*, starting with the right column and moving to the left; in case no digit appears in a certain spot, assume that the digit is 0.*

Schematically then for $4502 + 263$:

$$\begin{array}{rcccc} & 4 & 5 & 0 & 2 \\ + & & 2 & 6 & 3 \\ \hline & 4 & 7 & 6 & 5 \end{array} \qquad (12)$$

Thus starting from the ones digit (right column), we have: $2+3 = 5$, $0+6 = 6$, $5 + 2 = 7$, and since there is no digit in the spot below the number 4, the algorithm calls for putting a 0 there and the sum is $4+0 = 4$ for that column. Observe how the addition algorithm illustrates the *leitmotif* mentioned near the beginning of the section: the addition $4205 + 263$ is reduced to the calculation of four single-digit additions: $4 + 0$, $5 + 2$, $0 + 6$, and $2 + 3$.

The following discussion of addition will be restricted to this particular version of the algorithm. In due course, we shall discuss why the algorithm moves from right to left, which some people consider unnatural.

First of all, when this algorithm is taught in the classroom, the main emphasis is usually not on simple cases such as $4502 + 263$ where the sum of the digits in each column remains a single-digit number, but on cases such as $69 + 73$ where the process of "carrying" to the next column takes place. In this monograph, we reverse the emphasis by spending more time on the simple case before dealing with the more complicated case. There is a good reason for this decision: it is in the simple case that one gets to see with unobstructed clarity the main line of the logical reasoning, and when that is well understood, the more complicated case — which is nothing but the simple case embellished with a particular technique — tends to follow easily.

To underscore the fact that each step of the algorithm is strictly limited to the consideration of a single digit without regard to its place value, consider $865 + 32$:

$$\begin{array}{rccc} & 8 & 6 & 5 \\ + & & 3 & 2 \\ \hline & 8 & 9 & 7 \end{array} \tag{13}$$

Notice that the third column from the right of (12) and the rightmost column of (13) are identical:

$$\begin{array}{rc} & 5 \\ + & 2 \\ \hline & 7 \end{array} \tag{14}$$

Yet, in terms of place value, we know that in the context of (12), the addition fact in (14) actually stands for

$$\begin{array}{rccc} & 5 & 0 & 0 \\ + & 2 & 0 & 0 \\ \hline & 7 & 0 & 0 \end{array}$$

because the 5 in 4502 stands for 500 and the 2 in 263 stands for 200. By contrast, in the context of (13), the addition fact in (14) is literally true: it is just $2+5=7$. *As far as the algorithm is concerned*, however, the addition fact (14) is carried out in exactly the same way in (12) or (13), without regard to this difference.

Would this digit-by-digit feature of the algorithm corrupt students' understanding of place value? Not if students are made to understand that, far from a defect, this digit-by-digit feature is a virtue *for the purpose of easy computation*. Had the procedure of the algorithm treated each digit differently according to its place value, the algorithm would lose much of its simplicity: imagine that for the ones place you do one thing, for the tens place you do another, and for the hundreds place you do yet another, etc. How efficient can the algorithm be in that case? Moreover, in learning the *procedural* aspect of the algorithm, students should at the same time achieve the understanding that the algorithm is correct precisely because of place value considerations. To see this, recall:

$$\begin{aligned} 865 &= 8 \times 10^2 + 6 \times 10^1 + 5 \times 10^0 \\ 32 &= 0 \times 10^2 + 3 \times 10^1 + 2 \times 10^0 \end{aligned} \tag{15}$$

(Notice that we have made use of the trivial fact $32 = 032$ mentioned at the end of §1.) Recall an earlier comment made also in §1 about the occasional advantage of explicitly writing down all the powers of 10 in the expanded form of a number. We will see how this more clumsy notation lends conceptual clarity to what we have to do. We add these two equations. The left sides add up to $865 + 32$, of course. What about the right sides? We can add "vertically":

$$\begin{aligned} 8 \times 10^2 + 0 \times 10^2 &= (8+0) \times 10^2 \quad (= \; 8 \times 10^2) \\ 6 \times 10^1 + 3 \times 10^1 &= (6+3) \times 10^1 \quad (= \; 9 \times 10^1) \\ 5 \times 10^0 + 2 \times 10^0 &= (5+2) \times 10^0 \quad (= \; 7 \times 10^0)\,, \end{aligned}$$

where we have made use of the distributive law (10) three times in succession. (It may be instructive for you to read the exhortation near the end of §2 about the need to know that $MN + ML = M(N + L)$. Because the left side of the sum of equations equals the right side of this sum, we get

$$\begin{aligned} 865 + 32 &= (8+0) \times 10^2 + (6+3) \times 10^1 + (5+2) \times 10^0 \\ &(= 8 \times 10^2 + 9 \times 10^1 + 7 \times 10^0 \\ &= 897\ )\,, \end{aligned} \tag{16}$$

which is seen to be a parallel description of the addition algorithm in (13). We now see that

> *the addition algorithm is the method of adding two numbers by adding the digits corresponding to the same power of 10 when the two numbers are written out in their expanded forms.*

In particular, we see why the algorithm calls for replacing the empty spot under 8 with 0: it is none other than an abbreviated statement of $8 + 0 = 8$ in (16). The reasoning is valid in general and is by no means restricted to this special case of $865 + 32$.

Activity: Give a similar explanation of (12).

The preceding explanation of the addition algorithm — by this we mean the main ideas but not necessarily the precise notational formalism — would be adequate in most classroom situations. It is important to realize, however, that there are subtle gaps in the reasoning above, so that in the interest of a *complete* understanding, we proceed to fill in these gaps.

From (15), we have:

$$\begin{aligned} 865 + 32 &= \{8 \times 10^2 + 6 \times 10^1 + 5 \times 10^0\} + \\ &\quad \{0 \times 10^2 + 3 \times 10^1 + 2 \times 10^0\} \end{aligned}$$

As noted in connection with the associative and commutative laws of addition (5) and (6), we can ignore the braces and rearrange the order of summation of these six terms. Thus,

$$\begin{aligned} 865 + 32 &= \{8 \times 10^2 + 0 \times 10^2\} + \{6 \times 10^1 + 3 \times 10^1\} + \\ &\quad \{5 \times 10^0 + 2 \times 10^0\} \end{aligned}$$

It is at this point that we can apply the distributive law (10) to conclude (16). Thus what we said above about "adding vertically" in (15) is in fact an application in disguise of the associative and commutative laws of addition.

Pedagogical Comments: Should one emphasize discussion of the above type in an elementary school classroom? Few questions in education can be answered with absolute certainty, but there are at least two reasons why an *elaborate* discussion of the associative law and commutative law in, say, K–5 might interfere with a good mathematics education. First, such explanations tend to be somewhat tedious and students might lose interest and, second, such details might obscure the main thrust of the argument which is encapsulated in (16). As a teacher, however, you owe it to yourself to understand this kind of details. This is because, on the one hand, intellectual honesty demands it and, on the other, you must be prepared in case a precocious youngster presses you for the complete explanation. **End of Pedagogical Comments.**

For the case of $865 + 32$, it is possible to give a naive explanation of the addition algorithm in terms of money. Imagine that someone has

8 hundred-dollar bills,
6 ten-dollar bills, and
5 one-dollar bills,

thus \$865 altogether. Later she acquires another stack of bills consisting of

3	ten-dollar bills, and
2	one-dollar bills,

thus another \$32. To find out how much money she has altogether, she decides on the following strategy: collect all the hundred-dollar bills, then collect all the ten-dollar bills, and finally collect all the one-dollar bills. She finds that she now has

$8\ (= 8 + 0)$	hundred dollar bills
$9\ (= 6 + 3)$	ten-dollar bills
$7\ (= 5 + 2)$	one-dollar bills

So she has \$897. Exactly as in (13).

Before proceeding further, the question must be raised as to why it is *not* sufficient to understand the addition algorithm in terms of money alone, and why we must go through the previous mathematical explanation. The most superficial answer is that, because the algorithm is about numbers and not specifically about money, we should be able to offer an explanation that is valid in all contexts besides money. For example, why does this explanation also explain the fact that 865 crabs + 32 crabs = 897 crabs, or 865 stars + 32 stars = 897 stars ? By the time you have found the answer, it is most likely the case that you have also found a purely mathematical explanation along the line of (16). On a deeper level, we have just seen how the mathematical explanation brings out issues that are hidden in the explanation using money, such as place value and the basic laws of operations of §2 . The mathematical explanation brings a deeper understanding not only of the algorithm itself but also of these related issues, as we begin to see the interconnectedness of these seemingly disparate concepts. Moreover, we want an explanation that is sufficiently robust to be applicable to all numbers big or small. In this regard, it would be extremely clumsy to explain the addition $50,060,001 + 870,040$ in terms of money, but relatively easy to do so by the mathematical method above, as we shall see.

Let us now take up the issue of “carrying” from one column to the next. From our point of view, this issue is no more than a minor refinement of all that has been said before. Consider $68 + 59$:

$$\begin{array}{rccc} & & 6 & 8 \\ & & 5 & 9 \\ + & {\scriptstyle 1} & {\scriptstyle 1} & \\ \hline & 1 & 2 & 7 \end{array} \tag{17}$$

The difference from the previous situations is that, in the right column, the sum of digits, i.e., $8 + 9$, is no longer a single digit number but is 17. The algorithm calls for *carrying* the tens digit "1" of 17 to the next column on the left. This is indicated by the small 1 under 5, but in some other conventions, the 1 is entered above 6 instead; such minor notational differences are irrelevant to the mathematics under discussion. Then in adding the numbers in the tens column, this 1 is taken into account and result in the sum $6 + 5 + 1 = 12$, which is again not a single digit number. In like manner then, the "1" of 12 is carried to the (invisible) hundreds column. Because there is no other number in the hundreds column, the last 1 is recorded in the hundreds column and we get 127 as the final sum.

The basic explanation is the same as before, so it suffices to consider only the new features here. Let us begin with an explanation in terms of money. Imagine that you have a stack of bills consisting of

$$\begin{array}{cl} 6 & \text{ten-dollar bills, and} \\ 8 & \text{one-dollar bills,} \end{array}$$

and another stack consisting of

$$\begin{array}{cl} 5 & \text{ten-dollar bills, and} \\ 9 & \text{one-dollar bills.} \end{array}$$

You decide to count the two stacks by counting the ten-dollar bills and the one-dollar bills separately. Thus you have

$$\begin{array}{rl} 11\ (= 6 + 5) & \text{ten-dollar bills, and} \\ 17\ (= 8 + 9) & \text{one-dollar bills.} \end{array}$$

But 17 one-dollar bills is the same as 1 ten-dollar bill plus 7 one-dollar bills. So you trade 10 of your one-dollar bills for a ten-dollar bill and you now have

$$\left.\begin{array}{rl} 12\ (= 11 + 1) & \text{ten-dollar bills, and} \\ 7 & \text{one-dollar bills.} \end{array}\right\} \tag{18}$$

But we can also trade in 10 of the 12 ten-dollar bills for 1 hundred dollar bill, thereby getting

$$\left.\begin{array}{ll} 1 & \text{hundred-dollar bill,} \\ 2 & \text{ten-dollar bills, and} \\ 7 & \text{one-dollar bills.} \end{array}\right\} \qquad (19)$$

This line of reasoning not only arrives at the same answer, which is 127, but also exhibits the fact that (18) corresponds exactly to carrying the 1 to the tens place in (17), while (19) corresponds to carrying the 1 to the hundreds place in (17).

Finally, we give a purely mathematical explanation of (17). We preface the explanation with the comment that, having given the detailed reasoning in supported of the computation (13) citing the associative, commutative, and distributive laws, we will not overwhelm you with such details from now on but will instead ask you to be aware of their implicit presence at every turn. With this understood,

$$\begin{aligned} 68 + 59 &= (6 \times 10^1 + 8 \times 10^0) + (5 \times 10^1 + 9 \times 10^0) \\ &= (6 \times 10^1 + 5 \times 10^1) + (8 \times 10^0 + 9 \times 10^0) \end{aligned}$$

Applying the distributive law, we get

$$\begin{aligned} 68 + 59 &= (6 + 5) \times 10^1 + (8 + 9) \times 10^0 \\ &= (10 + 1) \times 10^1 + (10 + 7) \times 10^0 \\ &= (1 \times 10^2) + (1 \times 10^1) + (1 \times 10^1) + (7 \times 10^0) \\ &= (1 \times 10^2) + ([1 + 1] \times 10^1) + (7 \times 10^0). \end{aligned}$$

the last line explicitly shows how the 1 is carried to the next column to the left, twice, in (16). So $68 + 59 = 127$.

This is the right time and place to tie up a loose end. It was mentioned after (12) that there is a reason for the algorithm to move from right to left. The reason is one of *economy of means*: we want to simplify the algorithm to the utmost. Consider the preceding problem of $68 + 59$. Suppose we want to embark on a *different* addition algorithm by starting instead with the leftmost column and move to the right. Then the first step would be:

$$\begin{array}{rccc} & & 6 & 8 \\ + & & 5 & 9 \\ \hline & 1 & 1 & \end{array}$$

Here we have carried the 1 to the hundreds column. In the next step, we add the right column and get 17. This can be recorded in the next row again in accordance with the place value of 17 thus:

$$\begin{array}{cccc} & & 6 & 8 \\ + & & 5 & 9 \\ \hline & 1 & 1 & \\ & & 1 & 7 \end{array}$$

Then we add these two numbers and arrive at 127 as before:

$$\begin{array}{cccc} & & 6 & 8 \\ + & & 5 & 9 \\ \hline & 1 & 1 & \\ & & 1 & 7 \\ \hline & 1 & 2 & 7 \end{array}$$

The first comment is that *this algorithm is certainly correct*. Moreover, the proper execution of this algorithm requires the same technique of carrying the 1 to the next column, the only difference being that, whereas the carrying in (17) is recorded under 59, here it is done by using two extra rows. Comparing the last computation with (17) reveals one virtue of (17): it is compact, and this is the decisive factor. A beginner might welcome the left-to-right algorithm at the beginning, but once familiarity sets in, (17) is the algorithm of choice for most people. All the more so when the addition involves numbers with many digits.

A passing comment should be made about the "unnaturalness" of making children do things from right to left. One must put things in perspective. The fact is that children learn to do many things that *they* consider unnatural. Brushing teeth is one of them, and staying neat and clean is another. Moreover, inducing violent actions on a video screen by manipulating a joystick is quite possibly one of the most unnatural things imaginable for children. Yet, children all over the world welcome video or computer games with open arms. Are we then to believe that they cannot learn to calculate from right to left even after we explain to them why it is important that they do?

We have thus far discussed the addition algorithm in terms of concrete numbers for the simple reason that it is very difficult, notationwise, to discuss this and other whole number algorithms using abstract notation. Nevertheless, we hope that the generality of the underlying mathematical reasoning comes through so that, faced with other situations, you can both execute this algorithm correctly *and* know how to justify it. Just to be sure, we will discuss two more examples for further illustration. First, consider $165+27+83+829$. (Recall that this makes sense without parentheses because of the associative law.) Here is how the algorithm works out:

$$\begin{array}{rcccc} & & 1 & 6 & 5 \\ & & & 2 & 7 \\ & & & 8 & 3 \\ & & 8 & 2 & 9 \\ + & {\scriptstyle 1} & {\scriptstyle 2} & {\scriptstyle 2} & \\ \hline & 1 & 1 & 0 & 4 \end{array} \tag{20}$$

The precise description of the procedure is nothing new: *again we add column-by-column, and move from right to left.* Start with the right column and add: $5 + 7 + 3 + 9 = 24$, so we carry the 2 to the next column (to the left). Next, $(6 + 2 + 8 + 2) + 2 = 20$, so we again carry the 2 to the next column. Finally, in the hundreds column: $(1 + 8) + 2 = 11$, and we carry the 1 to the thousands column. Because there is no other number in the thousands column, we record the 1 in the final sum.

The explanation is the following:

$$\begin{aligned} 165 + 27 + 83 + 829 &= (1 \times 10^2) + (6 \times 10^1) + (5 \times 10^0) + \\ &\qquad (2 \times 10^1) + (7 \times 10^0) + \\ &\qquad (8 \times 10^1) + (3 \times 10^0) + \\ &\quad (8 \times 10^2) + (2 \times 10^1) + (9 \times 10^0) \\ &= (9 \times 10^2) + (18 \times 10^1) + (24 \times 10^0) \end{aligned}$$

But $24 \times 10^0 = (2 \times 10^1) + (4 \times 10^0)$, so by the distributive law:

$$\begin{aligned} 165 + 27 + 83 + 829 &= (9 \times 10^2) + (18 \times 10^1) + (2 \times 10^1) + (4 \times 10^0) \\ &= (9 \times 10^2) + ([18 + 2] \times 10^1) + (4 \times 10^0) \end{aligned}$$

This explains why we carried the 2 to the tens column in (20). Next, $[18 + 2] \times 10^1 = 2 \times 10^2$. Hence,

$$\begin{aligned} 165 + 27 + 83 + 829 &= (9 \times 10^2) + (2 \times 10^2) + (4 \times 10^0) \\ &= ([9 + 2] \times 10^2) + (4 \times 10^0) \end{aligned}$$

This explains why we carried the 2 to the hundreds column in (20). In any case, it is now clear that because $[9 + 2] \times 10^2 = [10 + 1] \times 10^2 =$

$(1 \times 10^3) + (1 \times 10^2)$, we obtain:

$$165 + 27 + 83 + 829 = (1 \times 10^3) + (1 \times 10^2) + (4 \times 10^0) = 1104\,,$$

exactly as in (20).

As a final example, let us use the addition algorithm to compute $50,060,001 + 870,040$ and at the same time give the explanation.

$$\begin{array}{rcccccccc} & 5 & 0 & 0 & 6 & 0 & 0 & 0 & 1 \\ & & & 8 & 7 & 0 & 0 & 4 & 0 \\ + & & & {\scriptstyle 1} & & & & & \\ \hline & 5 & 0 & 9 & 3 & 0 & 0 & 4 & 1 \end{array}$$

The explanation is simple:

$$\begin{aligned} 50,060,001 + 870,040 &= \{5 \times 10^7 + 6 \times 10^4 + 1 \times 10^0\} + \\ &\quad \{8 \times 10^5 + 7 \times 10^4 + 4 \times 10^1\} \\ &= (5 \times 10^7) + (8 \times 10^5) + ([6+7] \times 10^4) + \\ &\quad (4 \times 10^1) + (1 \times 10^0) \\ &= (5 \times 10^7) + (8 \times 10^5) + \\ &\quad ([10+3] \times 10^4) + (4 \times 10^1) + (1 \times 10^0) \end{aligned}$$

Look at the second and third terms of the right side: $(8 \times 10^5) + ([10+3] \times 10^4) = ([8+1] \times 10^5) + (3 \times 10^4)$. This then accounts for carrying the 1 to the column containing 8. Now, we obtain:

$$\begin{aligned} &50,060,001 + 870,040 = \\ &\qquad (5 \times 10^7) + (9 \times 10^5) + (3 \times 10^4) + (4 \times 10^1) + (1 \times 10^0)\,. \end{aligned}$$

As we mentioned above, an explanation of this addition problem using money would be both clumsy and irrelevant.

Exercise 3.1 Explain to a fourth grader why the addition algorithm for $57032 + 2845$ is correct, first using the method of (16), then using money.

Exercise 3.2 Explain to a fourth grader why the addition algorithm $826 + 4907$ is correct, with and without the use of money.

Exercise 3.3 Do the addition $67579 + 84937$ both ways, from left to right, and from right to left, and compare.

Exercise 3.4 Compute $123 + 69 + 528 + 4$ by the addition algorithm, and give an explanation of why the computation is correct.

Exercise 3.5 Compute $7826 + 7826 + 7826$ by the addition algorithm, and give an explanation of why the computation is correct.

Exercise 3.6 Compute $670309 + 95000874$ by the addition algorithm, and give an explanation of why the computation is correct.

3.2 A subtraction algorithm

Subtraction has to be understood in terms of addition. Thus $37 - 19$ is by definition the number so that, when added to 19, yields 37. Thus:

$$(37 - 19) + 19 = 37.$$

So *adding* 19 undoes the effect of *subtracting* 19. Similarly, if m and n are whole numbers, and $m < n$, then $n - m$ is *by definition* the number so that, when added to m, yields n. Thus,

$$(n - m) + m = n \qquad (21)$$

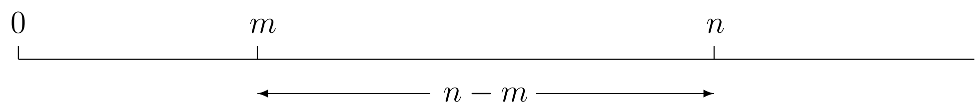
So again, *adding* m undoes the effect of *subtracting* m. It also means that in order to check whether a number x is equal to $n - m$, it is the same as checking whether

$$x + m = n.$$

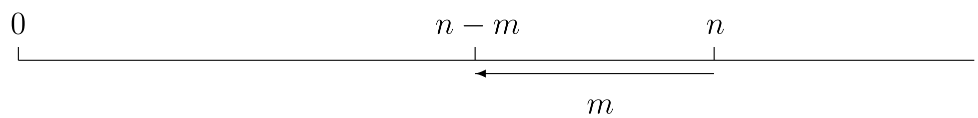
In particular, *to verify a statement about subtraction, it suffices to verify a statement about addition.* This simple fact will be seen to be extremely useful. On an intuitive level, it is common to think of $37 - 19$ as "taking 19 objects from 37 of them", and $n - m$ as "taking m objects from n of them".

> **Activity:** $1200 - 500 = ?$ $580,000,000 - 500,000,000 = ?$ $580,000,000 - 20,000,000 = ?$ $15 \times 10^6 - 7 \times 10^6 = ?$

In terms of the number line, $n - m$ has the following geometric interpretation. Rewrite (21) as $m + (n - m) = n$ (the commutativity of addition!). Note that m and n are the lengths of the segments $[0, m]$ and $[0, n]$. Therefore, according to the interpretation of the sum of two whole numbers as the length of a concatenated segment, the preceding equality implies that $n - m$ is exactly the length of the segment from m to n:



We can also look at (21) as is: $(n - m) + m = n$. This says that the concatenation of $[0, n - m]$ and $[0, m]$ is a segment of the same length as $[0, n]$. This means that we can arrive at the point $n - m$ by going from n *to the left* for a length of m:



As is the case with the addition algorithm, the purpose of the standard subtraction algorithm is to relieve the tedium of counting backwards 257 times from 658 in order to compute $658 - 257$. The mathematics underlying this algorithm (to be introduced presently) is so similar to that of the addition algorithm that we can afford to be brief. As before, we begin with the simple case, e.g., $658 - 257$, where the simplicity refers to the fact that each of the digits in the first number is at least as big as the corresponding digit in the second number. The algorithm calls for lining up the digits of the two numbers column-by-column from the right (as in the addition algorithm) and then do subtraction of single digit numbers in each column, starting with the right column and move left: $8 - 7 = 1$, $5 - 5 = 0$, $6 - 2 = 4$.[10] Schematically:

$$\begin{array}{rccc} & 6 & 5 & 8 \\ - & 2 & 5 & 7 \\ \hline & 4 & 0 & 1 \end{array} \qquad (22)$$

For the explanation of the algorithm, the following subtraction facts are useful. Suppose l, m, n, a, b, and c are any whole numbers so that $l - a$, $m > b$ and $n > c$, we have:

$$(l + m + n) - (a + b + c) = (l - a) + (m - b) + (n - c) \qquad (23)$$

$$m(l - a) = ml - ma \qquad (24)$$

[10] Again, the subtraction of $658 - 257$ is reduced to three single-digit computations: $6 - 2$, $5 - 5$, and $8 - 7$.

Equation (23) is entirely plausible if it is interpreted in terms of concrete objects. For example, if you have three piles of oranges, having l, m and n oranges in each pile, respectively, then taking $a+b+c$ oranges away from the three piles combined would leave behind the same total number of oranges as taking successively a oranges from the pile of l oranges, b oranges from the pile of m oranges, and c from the pile of n oranges. Equation (24) is a variant of the distributive law (10) and can be made believable using oranges in exactly the same way.

> It is also instructive to directly prove (23) and (24), because we will get to know the deeper meaning of the definition (21) in the process. First note that (23) makes sense, i.e., it is in fact true that $(l+m+n) > (a+b+c)$ so that the left side of (23) makes sense. To see why, we make repeated use of the third assertion of (11) near the end of §2 (to the effect that $A > B$ and $C > D$ imply $A+C > B+D$) and use the assumption of $l > a$, $m > b$, and $n > c$ to conclude that $(l+m+n) > (a+b+c)$. Now to prove (23), let $x = l - a$, $y = m - b$, and $z = n - c$, then the right side of (23) becomes $x+y+z$, and we want to prove that
>
> $$(l+m+n) - (a+b+c) = x+y+z.$$
>
> According to the remark after definition (21), this is the same as checking
>
> $$((a+b+c) + (x+y+z) = (l+m+n) \qquad (\clubsuit)$$
>
> By the general associative law and general commutative law of addition, the left side of (♣) is equal to $(x+a)+(y+b)+(z+c)$, which can be further simplified by use of
>
> $$\begin{array}{rcccl} x+a &=& (l-a)+a &=& l \\ y+b &=& (m-b)+b &=& m \\ z+c &=& (n-c)+c &=& n \end{array}$$
>
> Thus the left side of (♣) is $l+m+n$, which shows that (♣) is true.
>
> Simlarly, to show that (24) is true, let $x = l - a$. Then (24) becomes the statement that $mx = ml - ma$. Again by the remark after (21), this would be true if we can show
>
> $$mx + ma = ml \qquad (\spadesuit)$$
>
> But by the distributive law, the left side of (♠) is $m(x+a)$ which in virtue of $x = l - a$ is equal to $m([l-a]+a) = ml$. This shows that (♠) is also true. We have thus completed the formal proofs of (23) and (24)

It should be remarked that (23) is valid for more than three pairs of numbers. For example, the analogue of (23) for five pairs of numbers would read: if $l > a$, $m > b$, $n > c$, $p > d$, and $q > e$, then

$$
\begin{aligned}
&(l+m+n+p+q)-(a+b+c+d+e) = \\
&\quad (l-a)+(m-b)+(n-c)+(p-d)+(q-e)
\end{aligned}
$$

The proof of this more general version is of course the same as the case of three pairs of numbers. A complete understanding of (23) and (24) must await the introduction of negative numbers in Chapter 5.

Using (23) and (24), we can now give the explanation of (22):

$$
\begin{aligned}
658-257 &= (6\times 10^2+5\times 10^1+8\times 10^0)- \\
&\quad (2\times 10^2+5\times 10^1+7\times 10^0) \\
&= (6\times 10^2-2\times 10^2)+(5\times 10^1-5\times 10^1)+ \\
&\quad (8\times 10^0-7\times 10^0) \qquad \text{(by (23))} \\
&= ([6-2]\times 10^2)+([5-5]\times 10^1)+([8-7]\times 10^0) \\
&\qquad \text{(by (24)} \\
&= 4\times 10^2+1\times 10^0 \;=\; 401
\end{aligned}
$$

We now tackle the general case which requires "trading". Consider, for example, $756-389$. The preceding column-by-column method breaks down at the first step because the subtraction $6-9$ cannot be done *as is* using whole numbers. In this case, the algorithm states:

> *Take 1 from the tens digit (which is 5) of 756 — thereby changing the tens digit of 756 from 5 to 4 — and then do the subtraction in the ones column, not as* $6-9$*, but as* $16-9$ *(which is 7). In the tens column, we now have* $4-8$*. Again, take 1 from the hundreds digit (which is 7) of 756 — so that 7 becomes 6 — and do the subtraction in the tens column as* $14-8$ *(which is 6). Finally, the subtraction in the hundreds column is now* $6-3=3$*.*

Schematically:

$$
\begin{array}{rccc}
 & 6 & 4 & \\
 & \not{7} & \not{5} & 6 \\
- & 3 & 8 & 9 \\
\hline
 & 3 & 6 & 7
\end{array}
\qquad (25)
$$

The explanation for (25) and the algorithm itself is the reverse of the explanation given for carrying in the addition algorithm. We first do it schematically (note: the double arrow “ $\Longleftrightarrow$ ” in the following means “is the same as”):

$$\begin{array}{rccc} & 7 & 5 & 6 \\ - & 3 & 8 & 9 \\ \hline & ? & ? & ? \end{array} \quad\Longleftrightarrow\quad \begin{array}{rrrr} & 700+ & 50+ & 6 \\ - & 300+ & 80+ & 9 \\ \hline & & ? & \end{array}$$

$$\Longleftrightarrow\quad \begin{array}{rrrr} & 700+ & 40+ & 16 \\ - & 300+ & 80+ & 9 \\ \hline & & ? & \end{array} \quad\Longleftrightarrow\quad \begin{array}{rrrr} & 600+ & 140+ & 16 \\ - & 300+ & 80+ & 9 \\ \hline & & ? & \end{array}$$

$$\Longleftrightarrow\quad \begin{array}{rrrr} & 600+ & 140+ & 16 \\ - & 300+ & 80+ & 9 \\ \hline & 300+ & 60+ & 7 \end{array} \quad\Longleftrightarrow\quad \begin{array}{rccc} & 7 & 5 & 6 \\ - & 3 & 8 & 9 \\ \hline & 3 & 6 & 7 \end{array}$$

Now, in greater detail, we use the *associative law of addition repeatedly* to get:

$$\begin{aligned} 756 &= 700 + 50 + 6 \\ &= (600 + 100) + 50 + 6 \\ &= 600 + (100 + 50) + 6 \\ &= 600 + 150 + 6 \\ &= 600 + (140 + 10) + 6 \\ &= 600 + 140 + (10 + 6) \\ &= 600 + 140 + 16 \end{aligned}$$

so that

$$\begin{aligned} 756 - 389 &= (600 + 140 + 16) - (300 + 80 + 9) \\ &= (600 - 300) + (140 - 80) + (16 - 9) \qquad \text{(by (23))} \\ &= 300 + 60 + 7 \\ &= 367 \end{aligned}$$

Note that the second line of the preceding calculation corresponds exactly to the subtraction in (25). Note also that we avoided writing the number 756 and 389 in their expanded forms, but used a simplified version instead, because *the more complicated notation would have obscured the underlying reasoning.*

Although it is not a good policy in general to over-emphasize the role of the laws of operations discussed in §2, it is nevertheless worthwhile to point out the critical role played by the associative law of addition in the subtraction algorithm.

We add the usual remark that the mathematical reasoning behind the preceding explanation is perfectly general and is applicable to any subtraction $a-b$ for any whole numbers a and b, so long as a is not smaller than b. Just to be sure that this message gets across, we will work out another example: $5003 - 465$.

As before, the first subtraction in the right column, $3 - 5$, cannot be carried out using whole numbers. We try to take a 1 from the tens digit of 5003, which is unfortunately zero. So we go to the hundreds digit and hope to take 1 from there. Again it is 0. This then requires going all the way to the thousands digit "5" of 5003 to take 1 from 5. So 5 becomes 4 in the thousands digit of 5003, but the hundreds digit of 5003 becomes $10 + 0 = 10$. From this 10, we can take 1 to bring it down to the tens digit. in the process, the hundreds digit becomes 9 (instead of 10), and the tens digit is now $10+0 = 10$. We are now back to our original problem of getting 1 from the tens digit to change the 3 to $10 + 3 = 13$. We do so, and the tens digit becomes 9. So now, the subtraction in the ones column becomes $13 - 5 = 8$. The subtraction in the tens column becomes $9 - 6 = 3$, and that in the hundreds digit becomes $9 - 4 = 5$. The thousands digit is of course 4. Schematically:

$$\begin{array}{ccccc} & {\scriptstyle 4} & {\scriptstyle 9} & {\scriptstyle 9} & \\ & \not 5 & \not 0 & \not 0 & 3 \\ - & & 4 & 6 & 5 \\ \hline & 4 & 5 & 3 & 8 \end{array}$$

The explanation using the associative law of addition follows (this time, we will dispense with the use of parentheses altogether):

$$\begin{aligned} 5003 &= 5000 + 3 \\ &= 4000 + 1000 + 3 \\ &= 4000 + 900 + 100 + 3 \\ &= 4000 + 900 + 90 + 10 + 3 \\ &= 4000 + 900 + 90 + 13 \end{aligned}$$

So using the generalized version of (23) for four pairs of numbers, we have:

$$\begin{aligned} 5003 - 465 &= (4000 + 900 + 90 + 13) - (0 + 400 + 60 + 5) \\ &= (4000 - 0) + (900 - 400) + (90 - 60) + (13 - 5) \\ &= 4000 + 500 + 30 + 8 \\ &= 4538 \end{aligned}$$

We note that the subtraction algorithm is again one that works from right to left. Just as in the case of the addition algorithm, one can work from left to right, but the amount of backtracking needed for making corrections is even greater here than in the case of addition.

> **Activity:** Do the preceding subtraction $5003 - 465$ from left to right, and compare with the computation from right to left.

It is worth repeating that there is absolutely nothing unnatural about teaching children to do something from right to left.

For special numbers, there are usually tricks to make computations with them much more pleasant. This applies in particular to the subtraction algorithm. Let us give one such example: the preceding problem $5003 - 465$ can be done very simply as follows:

$$5003 - 465 = 4 + 4999 - 465 = 4 + 4534 = 4538\,.$$

Similarly,

$$30024 - 8697 = 25 + 29999 - 8697 = 25 + 21302 = 21327.$$

The point here is that the subtractions with rows of 9's in the first number can be done without trading and can therefore be done easily by mental math. One can even do it from left to right if one wishes. The same trick can be used for any subtraction problem in which the first number is close to a multiple of 10^n, where n is any whole number. Then by changing the first number to a small number plus another one with a row of 9's, the subtraction can be done with no effort.

The preceding algorithm for subtraction when the first number is close to a multiple of 10^n is so striking that it would have spoiled the fun if we had called attention to a technical point. Nevertheless, mathematics must be served, if only discretely this time, so we attend to this technical point here. Take, for instance, $5003 - 465$. The point is that the step $4 + 4999 - 465 = 4 + 4534$ actually involves the associative law for integers (positive and negative whole numbers, see Chapter 5). Precisely, we have

$$\begin{aligned}(4+4999)-465 &= (4+4999)+(-465) \\ &= 4+(4999+(-465)) \\ &= 4+4534\end{aligned}$$

The reason why the preceding is worthy of a separate discussion is of course the fact that *the associative law in its naive form is not always valid for subtraction*, e.g., $(19 - 4) - 5 \neq 19 - (4 - 5)$. The mystery surrounding the associative law for subtraction would be dispelled once we do integer operations correctly. See Chapter 5.

As a final note on the subtraction algorithm, there is a natural alternative algorithm if negative numbers — together with a few simple arithmetic properties — are allowed to be used. To illustrate, consider the previous problem of $756 - 389$: *We continue to do column-by-column single digit subtractions*, but now we can use negative numbers to record the differences and we are free to do the column-by-column subtractions in any order, from left to right or right to left. So:

$$\begin{array}{rccc} & 7 & 5 & 6 \\ - & 3 & 8 & 9 \\ \hline & 4 & [\![-3]\!] & [\![-3]\!] \end{array}$$

where each $[\![-3]\!]$ indicates the result of the column subtraction *in that column*. To get the final answer, the algorithm says:

treat $4[\![-3]\!][\![-3]\!]$ *as if it were a whole number with digits 4,* $[\![-3]\!]$*, and* $[\![-3]\!]$*, and write it out in expanded form.*

Thus:

$$\begin{aligned} 4[\![-3]\!][\![-3]\!] &= (4 \times 10^2) + ((-3) \times 10^1) + ((-3) \times 10^0) \\ &= 400 - 30 - 3 \\ &= 370 - 3 \;=\; 367\,. \end{aligned}$$

Notice that the subtractions here are much more tractable than those in the original.

The explanation is simple enough and is based on (23) and (24):

$$\begin{aligned} 756 - 389 &= \{(7 \times 10^2) + (5 \times 10^1) + (6 \times 10^0)\} - \\ &\quad\ \{(3 \times 10^2) + (8 \times 10^1) + (9 \times 10^0)\} \\ &= ([7-3] \times 10^2) + ([5-8] \times 10^1) + ([6-9] \times 10^0) \\ &= 400 + ((-3) \times 10) + (-3) \;=\; 367 \end{aligned}$$

There are pros and cons regarding which of two algorithms is "better". The first is simpler, but the second may be less prone to computational errors, at least if the user is fluent with negative numbers. Because teaching third graders about the most elementary aspects of negative numbers is a realistic goal, the second algorithm should be a viable option in schools.

Exercise 3.7 Give an interpretation of (22) in terms of money.

Exercise 3.8 Explain to a fourth grader why the subtraction algorithm for $563 - 241$ is correct, with or without money.

Exercise 3.9 Give an interpretation of (25) in terms of money.

Exercise 3.10 Explain to a fourth grader why the subtraction algorithm for $627 - 488$ is correct, with and without the use of money.

Exercise 3.11 (a) Use the subtraction algorithm to compute $2403 - 876$ and explain why it is correct. (b) Do the same with $76431 - 58914$.

Exercise 3.12 Compute $800,400 - 770,982$ in two different ways, and explain why what you have done is correct.

Exercise 3.13 Compute $26,004 - 8325$ two ways, once using the standard algorithm, and once using the preceding "negative number" algorithm.

Exercise 3.14 Let a, b, c be whole numbers. (a) Prove that $a + b < c$ is the same as $a < c - b$. (b) Suppose $c < a$ and $c < b$. Prove that $a < b$ is the same as $a - c < b - c$.

Exercise 3.15 Find shortcuts to compute the following: $8 \times 875 = ?$ $9996 \times 25 = ?$ $103 \times 97 = ?$ $86 \times 94 = ?$

3.3 A multiplication algorithm

We next take up the question of how to compute the product of two numbers such as 826×73 *without* having to add 73 to itself 826 times. Bearing in mind the leitmotif enunciated at the beginning of the section, we proceed to break up the computation into a series of computations involving one digit at a time. In this case, the distributive law does the breaking up:

$$\begin{aligned} 826 \times 73 &= 826 \times (7 \times 10 + 3) \\ &= (\underline{826 \times 7}) \times 10 \;+\; (\underline{826 \times 3}) \end{aligned} \tag{26}$$

Thus

> *the multiplication (of 826) by a multi-digit number 73 has been reduced to two simpler computations: multiplication by 7 (i.e.,* 826×7*) and multiplication by 3* (826×3)*, as given by (26).*

We now further break up these two tasks into yet simpler tasks.

Let us first look at 826×3. Instead of adding 3 to itself 826 times, we apply the distributive law one more time:

$$\begin{aligned} 826 \times 3 &= \{(8 \times 10^2) + (2 \times 10) + 6\} \times 3 \\ &= (\underline{8 \times 3}) \times 10^2 + (\underline{2 \times 3}) \times 10 + (\underline{6 \times 3}) \end{aligned} \tag{27}$$

where of course we have also made use of the associative and commutative laws of multiplication to conclude that, e.g., $(8 \times 10^2) \times 3 = 8 \times 3 \times 10^2$. See the discussion in §2 above equation (10). Thus 826×3 will be computable according to (27), as soon as we know the products of single-digit numbers: 8×3, 2×3, and 6×3. **This is why a fluent knowledge of the multiplication table is essential,** because it lies at the heart of all multiplication problems. In any case, from (27) we obtain:

$$826 \times 3 = (24 \times 10^2) + (6 \times 10) + 18.$$

Experience with the addition and subtraction algorithms tells us that we should proceed by working from right to left: $18 = 10 + 8$, so that

$$\begin{aligned} 826 \times 3 &= (24 \times 10^2) + \ ([6+1] \times 10) \ + 8 \\ &= (24 \times 10^2) + (7 \times 10) + 8 \end{aligned} \tag{28}$$

Now $24 \times 10^2 = ([2 \times 10] + 4) \times 10^2 = (2 \times 10^3) + (4 \times 10^2)$, so that

$$826 \times 3 = (2 \times 10^3) + (4 \times 10^2) + (7 \times 10) + 8 \tag{29}$$

Equations (27)–(29) explain the following multiplication algorithm when one number is single-digit:

$$\begin{array}{rcccc} & & 8 & 2 & 6 \\ & & & & 3 \\ \times & {\scriptstyle 2} & & {\scriptstyle 1} & \\ \hline & 2 & 4 & 7 & 8 \end{array} \tag{30}$$

The precise description of the algorithm for the multiplication of 826 by a single digit number 3 is this. *Multiply each digit of 826 by 3, from right to left:* $3 \times 6 = 18$, *so carry the 1 to the tens column;* $3 \times 2 = 6$, *so the tens digit of the answer is* $6 + 1 = 7$*;* $3 \times 8 = 24$, *so carry the 2 to the hundreds column and get 24 for the thousands and hundreds columns respectively.*

To make sure that the multiplication algorithm with a single-digit multiplier is clearly understood, we will quickly do another example, 826×7, and give a brief explanation.

$$\begin{array}{rcccc} & & 8 & 2 & 6 \\ & & & & 7 \\ \times & {\scriptstyle 5} & {\scriptstyle 1} & {\scriptstyle 4} & \\ \hline & 5 & 7 & 8 & 2 \end{array}$$

It yields the right answer because:

$$\begin{aligned} 826 \times 7 &= \{(8 \times 10^2) + (2 \times 10) + 6\} \times 7 \\ &= (56 \times 10^2) + (14 \times 10) + 42 \\ &= (56 \times 10^2) + (14 \times 10) + ([4 \times 10] + 2) \\ &= (56 \times 10^2) + ([14 + 4] \times 10) + 2\,. \end{aligned}$$

The last line explains the carrying of the 4 to the tens column. Then:

$$\begin{aligned} 826 \times 7 &= (56 \times 10^2) + (18 \times 10) + 2 \\ &= (56 \times 10^2) + (10^2 + [8 \times 10]) + 2 \\ &= ([56 + 1] \times 10^2) + (8 \times 10) + 2\,, \end{aligned}$$

and the last line explains the carrying of the 1 to the hundreds column.

Recall that we are looking at the multiplication algorithm with a single-digit multiplier because (26) reduces the general case to this special case. It is time to return to the general case, say (26). We now know:

$$\begin{aligned} 826 \times 7 &= 5782 \\ 826 \times 3 &= 2478\,. \end{aligned}$$

According to (26) then,

$$\begin{aligned} 826 \times 73 &= (5782 \times 10) + 2478 \\ &= 2478 + 57820\,, \end{aligned}$$

so that schematically we have:

$$\begin{array}{rccccc} & & & 8 & 2 & 6 \\ \times & & & & 7 & 3 \\ \hline & & 2 & 4 & 7 & 8 \\ + & 5 & 7 & 8 & 2 & 0 \\ \hline & 6 & 0 & 2 & 9 & 8 \end{array}$$

Because we are used to treating an empty slot as a zero (cf. the discussion of the addition algorithm right above (12), for example), it is customary to omit the "0" at the end of 57820 and just write:

$$\begin{array}{rccccc} & & & 8 & 2 & 6 \\ \times & & & & 7 & 3 \\ \hline & & 2 & 4 & 7 & 8 \\ + & 5 & 7 & 8 & 2 & \\ \hline & 6 & 0 & 2 & 9 & 8 \end{array} \qquad (31)$$

We are in a position to summarize this algorithm as follows:

> *To compute say* 826 × 73*, take the digits of the second factor 73 individually, compute the two products with single digit multiplier — i.e.,* 826 × 3 *and* 826 × 7 *— and, when adding them, shift the one involving the tens digit (i.e., 7) one digit to the left, as in (31).*

This is commonly called the standard *multiplication algorithm*. Other variations are possible. For example, one alternative algorithm is

$$
\begin{array}{cccccc}
 & & & 8 & 2 & 6 \\
\times & & & & 7 & 3 \\
\hline
 & 5 & 7 & 8 & 2 & \\
+ & & 2 & 4 & 7 & 8 \\
\hline
 & 6 & 0 & 2 & 9 & 8
\end{array}
$$

Essentially, this means we run the algorithm from left to right, first multiply by 7 before mutiplying by 3. Mathematically, we do not consider such formal differences to be a difference at all. It is worth noting that even the algorithm with a single-digit multiplier can be carried out from left to right. For example, 6718×5 can be done this way:

$$
\begin{array}{cccccc}
 & & 6 & 7 & 1 & 8 \\
\times & & & & & 5 \\
\hline
 & 3 & 0 & & & \\
 & & 3 & 5 & & \\
 & & & & 5 & \\
+ & & & & 4 & 0 \\
\hline
 & 3 & 3 & 5 & 9 & 0
\end{array}
$$

Activity: Give a precise explanation of the preceding algorithm.

Pedagogical Comment: In a classroom, the most salient feature of this algorithm that catches students' attention may well be the shifting of the product involving the tens digit (i.e., 826×7) to the left by one digit. This should be carefully explained to them in terms of place value: we are actually looking at 826×70 and the shifting of digit is caused by the presence of the "0" in the ones digit. **End of Pedagogical Comment.**

In order to ensure that the generality of the preceding reasoning behind the algorithm is understood, let us briefly explain how to do a more complicated problem.

$$\begin{array}{rrrrrrr} & & & & 5 & 2 & 7 \\ \times & & & & 3 & 6 & 4 \\ \hline & & & 2 & 1 & 0 & 8 \\ & & 3 & 1 & 6 & 2 & \\ + & 1 & 5 & 8 & 1 & & \\ \hline & 1 & 9 & 1 & 8 & 2 & 8 \end{array} \qquad (32)$$

This is because:

$$\begin{aligned} 527 \times 4 &= 2108 \\ 527 \times 6 &= 3162 \\ 527 \times 3 &= 1582\,, \end{aligned}$$

and so by the distributive law:

$$\begin{aligned} 527 \times 364 &= 527 \times (3 \times 10^2 \ + \ 6 \times 10 \ + \ 4) \\ &= (\underline{527 \times 3}) \times 10^2 \ + \ (\underline{527 \times 6}) \times 10 \ + \ (\underline{527 \times 4}) \\ &= 158100 \ + \ 31620 \ + \ 2108 \end{aligned}$$

The last line explains the vertical alignments of the digits of the three products 2108, 3162, and 1581 in (32). In particular, because the 3 of 364 is the hundreds digit, the product 527×3 is shifted two digits to the left in keeping with the fact that it is really 527×300 that we are adding.

Exercise 3.16 Explain to a 4th grader why the multiplication algorithm for 86×37 is correct.

Exercise 3.17 (a) Which 2-digit number, when multiplied by 89, gives a 4-digit number that begins and ends with a 6? (b) List all the 3-digit numbers which have the following properties: the sum of the digits is 12, and when multiplied by 15 they give a 5-digit number which ends with a 5 (i.e., the ones digit is 5). (Clearly this problem can be done by guess and check. You are however asked to use reasoning to quickly dospatch it by narrowing down the choices.)

Exercise 3.18 Use the multiplication algorithm to compute

$$\begin{array}{rrrrrrr} & & & & & 1 & 8 \\ \times & 5 & 0 & 0 & 0 & 9 & 2 \\ \hline & & & ? & & & \end{array}$$

and explain why it is correct. Compare with the same algorithm applied to

$$\begin{array}{r} 5\ 0\ 0\ 0\ 9\ 2 \\ \times \qquad\qquad 1\ 8 \\ \hline ? \end{array}$$

Exercise 3.19 Compute 4208×879 by the multiplication algorithm and explain why it is correct.

3.4 Division-with-remainder

When asked to divide 23 by 4, we all know the answer: the quotient is 5 and the remainder is 3. In general, given whole numbers a and b with $b \neq 0$, we likewise want to know what the quotient and remainder are when a is divided by b. This knowledge is critical not only for the long division algorithm of §3.5 below but also for the discussions of fractions in Chapter 2 and decimals in Chapter 4. Before we can come up with an answer, however, we need to know precisely what the "quotient" of $a \div b$ means and what the "remainder" means. Most school textbooks do not deem it necessary to explain these concepts, or if they do, they typically say the following:

> *division* An operation on two numbers that tells how many groups or how many in each group.
>
> *quotient* The answer in division.
>
> *remainder* The number that is left over after dividing.

If you look more closely at the text proper, you would find the statement that the remainder should be less than the "divisor" b, but this still leaves out a clear statement about what a "quotient" is and what "left over after dividing" means. Such vagueness would not serve any purpose because unless the student already knows what "division" means, the above explanations give no information. For example, how do we compute the "quotient" of $6810255956001 \div 28747$ if we do not know its *precise* meaning? The usual attempt at an explanation of division would mention taking away multiples of 28747 until "the remainder" is "smaller than 28747". This unfortunately begs the question of what a "quotient" is and whether a negative number may be considered to be "smaller than" 28747.

As teachers, we want to convey the clear message to our children that in mathematics, no guesswork is needed for its mastery. We want to let them know that it is an open book that everybody can read. Among all branches of knowledge, mathematics is characterized by its WYSIWYG quality — what you see is what you get — and you have no need to assume anything that is not already explicitly stated. This is another way of expressing the fact that every conclusion we draw in mathematics depends completely on what is stated explictly up front. This is what we have been doing so far, and we intend to continue doing it for the rest of this monograph.

To return to division, let us first fix the meaning of $a \div b$ where a and b are whole numbers and $b \neq 0$. (We will always assume $b \neq 0$ in this situation because we do not want to divide by zero; see the discussion in §4 below.) Generally speaking, division is to multiplication as subtraction is to addition: one undoes the other. However, there are certain wrinkles to this simple-minded statement, and we will be careful to address them.

Recall that multiplication is repeated addition: by definition,

$$qb = \underbrace{b + b + \cdots + b}_{q}$$

(see (2) of §1). Division $a \div b$ is roughly speaking *repeated subtraction*, but the precise meaning of this phrase requires a rather long-winded explanation. First of all, if m is a whole number, the product mb is called a *multiple* of b, or more precisely, *the* m*-th multiple of* b. (In particular, 0 is a multiple of b, by definition because $0 = 0 \times b$.) Intuitively, what we are going to do is to repeatedly subtract b from a until we get to the point where the next subtraction will not be possible because what is left is smaller than b (recall: for $x - b$ to make sense, we must have $x \geq b$). Symbolically though, this way of doing things is very awkward, so we do something which is the same[11] but which is easier to express. In greater detail, what we do is to take (i.e., subtract) successive multiples of b from a: $a - 0$, $a - b$, $a - 2b$, $a - 3b$, $a - 4b$, . . . , until eventually we come to a multiple qb of b so that the next multiple — which is $(q + 1)b$ — exceeds a. In symbols, q is that whole number so that

$$a \geq qb \qquad \text{but} \qquad a < (q + 1)b. \tag{33}$$

[11] An explanation of why it is the same is given in the next fine-print indented passage.

By definition, we call this q the *quotient* of $a \div b$. It follows that the quotient q is the *largest* multiple of b that one can take away from a, because we cannot perform the next subtraction $a-(q+1)b$ for the reason that $a<(q+1)b$, by (33). It is intuitively clear, and we shall prove precisely below, that after we have taken q multiples of b's from a, what is left behind is less than b. We call $a-qb$ the *remainder* of $a \div b$. To complete the terminology, we call b the *divisor* and a the *dividend* of the division $a \div b$.

We now proceed to bring out a critical property of the remainder that was alluded to above, namely, $a-qb<b$. We first give a numerical proof, but presently we will also give a pictorial proof that makes the reasoning perfectly obvious. Here then is the numerical proof: from (33) we have $a<(q+1)b$. By taking qb away from both sides, the inequality does not change (see Exercise 3.14(b)). So we get

$$\begin{aligned} a-qb &< (q+1)b-qb \\ &= \{(q+1)-q\}b \qquad\qquad \text{(by (24) of §3.2)} \\ &= b \end{aligned}$$

So we get $a-qb<b$. Incidentally, we also know from (33) that $a \geq qb$, so $a-qb \geq 0$. We may therefore summarize the preceding two facts in the following *double inequality*:

$$\text{the remainder } a-qb \text{ satisfies} \qquad 0 \leq a-qb<b \tag{34}$$

Note that the left inequality sign of (34) is a weak inequality (i.e., allowing for equality) because the remainder does equal 0 sometimes, e.g., when $a=qb$.

For example, if $a=23$ and $b=4$, then the multiples of 4 are 4, $2\times 4=8$, $3\times 4=12$, $4\times 4=16$, $5\times 4=20$, $6\times 4=24$, and we stop at the 6th multiple of 4 because already

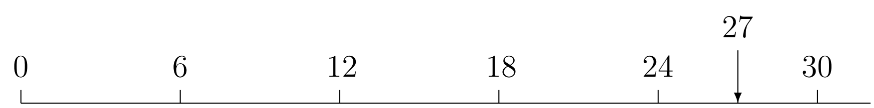
$$23 \geq (5\times 4) \qquad \text{but} \qquad 23<(6\times 4).$$

Therefore the quotient of $23 \div 4$ is 5, and the remainder is $23-(5\times 4)=3$. If we take $a=12$ and $b=3$, however, then we get the happy coincident that $12=4\times 3$, i.e., 12 is exactly the 4th multiple of 3. In this case, the quotient of $12 \div 3$ is 4, with remainder 0.

If $a=qb$ for some whole number q, we say b *divides* a. In symbols: $b|a$. Note that " b divides a" says exactly the same thing as " a is a multiple of b".

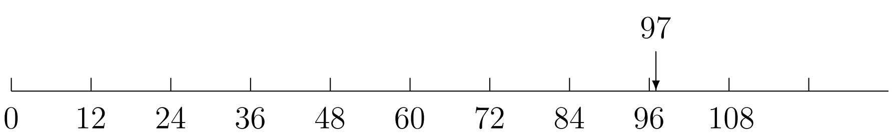
> **Activity:** Find the quotient and remainder in each of the following divisions by listing the multiples of the divisor (but do *not* use "long division", whatever that means): $33 \div 7$, $46 \div 9$, $98 \div 19$, $188 \div 37$.

We now give the pictorial representation of the quotient and remainder using the number line. Consider $27 \div 6$. The multiples of 6 are 0, 6, 12, 18, 24, 30, etc., as shown.



The picture clearly displays the fact that 27 is trapped between the two multiples of 6: 24 and 30. The remainder of $27 \div 6$ is just the length of the segment between 27 and $4 \times 6 = 24$, which is clearly less than the length between 24 and 30. Therefore the remainder $27 - (4 \times 6) < 6$ $(= 30 - 24)$. Thenquotient is 4 because $24 = 4 \times 6$.

As another example, consider $97 \div 12$. The multiples of 12 are: 0, 12, 24, 36, 48, 60, 72, 84, 96, 108.



Again, the dividend 97 is trapped between the two multiples 96 and 108 of 12, and the remainder of $97 \div 12$ is the length of the segment $[96, 97]$ (the segment between 96 and 97), which is less than 6 (= the length of the segment $[96, 108]$). Because $96 = 8 \times 12$, the quotient of $97 \div 12$ is 8.

In general, we have $a \div b$, then the multiples of b are equally spaced markers (= points) on the number line, b units apart. The whole number a has to be trapped between two of these multiples, or right at one of the multiples. In the former case, let a be between qb and $(q+1)b$.

Then, the remainder is just the length of the segment $[qb, a]$, and it is clearly smaller than the length of the segment $[qb, (q+1)b]$ $(= b)$.

On the other hand, if a is at one of the multiples of b, let it be at qb, as shown.

In this case, $a = qb$ and of course the remainder is 0.

We can summarize our discussion in the following theorem.

Division-with-Remainder. *Given any two whole numbers a and b, with $b > 0$, there exist a whole number q, called the* quotient *of $a \div b$, so that the* remainder *$a - qb$ satisfies* (34)*, i.e.,*

$$0 \le a - qb < b.$$

This theorem is more commonly cast in a different form, as follows. Let the remainder $a - qb$ be denoted by r, then $a - qb = r$ by definition, which can be rewritten as $a = qb + r$. The condition (34) now says $0 \le r < b$. Hence, we have an equivalent formulation of Division-with-Remainder:

Division-with-Remainder (Second Form). *Given any two whole numbers a and b, with $b > 0$, there exist a whole number q, called the* quotient *of $a \div b$, and a whole number r, called the* remainder *of $a \div b$, so that*

$$a = qb + r \qquad \text{where } r \text{ satisfies} \qquad 0 \le r < b. \tag{35}$$

An added remark about the division-with-remainder will be relevant in the discussion of long division (§3.5) and the decimal expansion of a fraction (§4 of Chapter 4). Ordinarily when one divides a by b, there is an implicit assumption that a is bigger than b. However, in the above statement of division-with-remainder, the relative sizes of a and b are irrelevant. For example, for $5 \div 32$, we have $5 = (0 \times 32) + 5$, so that the quotient is 0 and the remainder is 5. For $29 \div 127$, we have $29 = (0 \times 127) + 29$, with quotient 0 and remainder 29. The point is that the division-with-remainder makes sense for any $a \div b$, so long as $b > 0$.

We now tie up two loose ends left dangling in the preceding discussion. The first is the explanation of why "repeated subtraction of b from a" is the same as "subtracting successive multiples of b from a". The second is the the omission of the uniqueness of the quotient in the division-with-remainder.

Repeated subtractions of b from a means of course $(a-b)$, $((a-b)-b)$, $(((a-b)-b)-b)$, ... We used instead $a-b$, $a-2b$, $a-3b$, ..., in the above. What needs to be pointed out is that $((a-b)-b) = a-2b$, $(((a-b)-b)-b) = a-3b$, and in general

$$(\cdots((a-\underbrace{b)-b)\cdots-b}_{q}) = a-qb \qquad (\dagger)$$

for all whole numbers q. Of course (†) is intuitively obvious, because whether one takes b away from a one at a time, q times in succession, or take qb away from a all at once, what is left behind should be the same, However, it is also important to realize that intuition need not be the sole arbiter of mathematical truths, so we shall give a precise proof of (†) which also happens to be instructive.

We begin with an observation: if A, B are whole numbers so that $(a-M)-N \geq 0$, then

$$(a-M)-N = a-(M+N) \qquad (\ddagger)$$

Here is the reason: Let $x = (a-M)-N$. Then (‡) becomes the statement that $x = a-(M+N)$. Therefore, to prove (‡), we only need to prove $x = a-(M+N)$, and for this, — we recall the remark made after (21) in §3.2 — it suffices to prove $x+(M+N) = a$. To this end, observe again as a consequence of the definition of subtraction in (21) that $x = (a-M)-N$ means

$$x+N = a-M. \qquad (\star)$$

So we have:

$$\begin{aligned} x+(M+N) &= x+(N+M) && \text{(commutative law of +)} \\ &= (x+N)+M && \text{(associative law of +)} \\ &= (a-M)+M && \text{(by } (\star)) \\ &= a && \text{(by definition of subtraction in (21))} \end{aligned}$$

This proves $x+(M+N) = a$, and therewith (‡).

We can now prove (†) in succession for $q=1$, $q=2$, $q=3$, ..., etc. If $q=1$, (†) merely says $a-b = a-b$. Let $q=2$, then (†) states that

$$[a-b]-b = a-2b, \qquad (\diamond)$$

but this follows directly from (‡) by letting $M = N = b$. Next let $q=3$. Then we must prove:

$$([a-b]-b)-b = a-3b. \qquad (\diamond\diamond)$$

This is true because:

$$\begin{aligned}([a-b]-b)-b &= (a-2b)-b && \text{(by } (\diamond)\text{)}\\ &= a-(2b+b) && \text{(by } (\ddagger) \text{ with } M=2b \text{ and } n=b\text{)}\\ &= a-3b.\end{aligned}$$

Now let $q = 4$. Then we have to prove

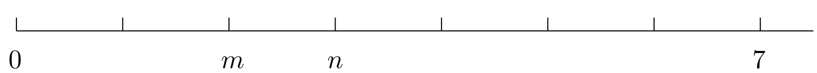
$$((([a-b]-b)-b)-b = a-4b,$$

and this is true because

$$\begin{aligned}((([a-b]-b)-b)-b &= (a-3b)-b && \text{(by } (\diamond\diamond)\text{)}\\ &= a-(3b+b) && \text{(by } (\ddagger) \text{ with } M=3b \text{ and } n=b\text{)}\\ &= a-4b,\end{aligned}$$

as desired. The next thing to prove is the case of (†) for $q = 5$, etc. But it is clear by now, with the pattern of proof firmly established, that rest of the argument would proceed would proceed in similar fashion. Therefore (†) is true for all whole numbers q.

We next turn to the division-with-remainder and point out that the quotient q of the theorem is actually *unique*. What this means is that suppose we have $31 \div 7$. Then we know the quotient is 4, and $0 \le 31-(4\times 7) < 7$. Now suppose there is another whole number s so that $0 \le 31-(s\times 7) < 7$. The theorem implies that this s must be equal to 4 too. The reason for this is based on a general fact: if we have two whole numbers m and n so that $0 \le m, n < 7$, — and we may as well assume $m \le n$ — then it is quite clear that the difference $n - m$ is a whole number which is not a nonzero multiple of 7. See the picture:



We now apply this simple observation to $31\div 7$ by letting $m = 31-(4\times 7)$ and $n = 31-(s\times 7)$. As before, we may assume $m \le n$. Then we know that $m-n$ is not a nonzero multiple of 7. Using the simple fact that $(l-a)-(m-b) = (l-m)+(b-a)$, to be proved in Chapter 5, we have:

$$\begin{aligned}m-n &= \{31-(4\times 7)\}-\{31-(s\times 7)\}\\ &= (31-31)+\{(s\times 7)-(4\times 7)\}\\ &= (s-4)\times 7\end{aligned}$$

Because we know ahead of time that $m - n$ is a whole number, we see that $(s-4)\times 7$ must be a whole number, and as such, it has to be a multiple of

7. We already know that it cannot be a *nonzero* multiple of 7, so it can only be the zero multiple, in which case $s - 4 = 0$, which is to say, $s = 4$. Exactly as claimed.

In general, the quotient q of $a \div b$ is unique in the sense that if s is another whole number so that also $0 \leq a - sb < b$, then necessarily $s = q$. The proof is identical to the preceding argument as soon as the numbers 31, 7 and 4 are repalced by a, b, and q, respectively. We will not repeat the argument. *Moreover, if the quotient is unique, then so is the remainder $a - qb$.*

Conceptually, the uniqueness is important because we talk freely about *the* quotient of a division and *the* remainder of a division. So without being aware of it, we tacitly assume the uniqness in question. More is true.

> *The requirement that both q and r be whole numbers is critical to their uniqueness.*

For example, the first example $23 \div 4$ of this subsection has (as we know) quotient and remainder equal to 5 and 3, respectively. However, if we are allowed to use fractions, for example, then we could have

$$23 = \boxed{5 + \frac{1}{4}} \times 4 + \boxed{2} \qquad \text{and} \qquad 0 \leq \boxed{2} < 4$$

(Although we have not yet taken up the subject of fractions, there is no harm in using them for illustration.) This would give a "quotient" of $5 + \frac{1}{4}$ and a "remainder" 2. In fact, we can write many such equations at will, e.g.,

$$23 = \boxed{\frac{22}{4}} \times 4 + \boxed{1} \qquad \text{and} \qquad 0 \leq \boxed{1} < 4$$

This should serve as reminder how delicate the uniqueness of the quotient and remainder really is. In school mathematics, however, such subtlety is usually glossed over. While one cannot say that such negligence does elementary school mathematics great harm, we hope nevertheless to have convinced you that, as a teacher, you should be aware of it.

The special case of division-with-remainder where the remainder is 0 occupies a place of distinction, and we proceed to discuss it in some detail. So let $a = qb$, where q is the quotient. In this case, it is customary to write $q = a \div b$. It follows that if $a = qb$, then

$$(a \div b) \times b = a \qquad \text{and} \qquad (ab) \div b = a \tag{36}$$

(It is of some value to point out that the second assertion is strictly a consequence of the definition: let $x = ab$, then by the definition above, $a = x \div b$, which is another way of writing $a = (ab) \div b$.) The equations in (36) clearly display division and multiplication as two operations that undoes each other.

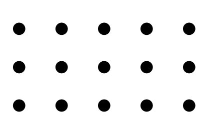
If $q = a \div b$, the fact that $a = qb = b + b + \cdots + b$ (q times) means that if we take b objects from a each time and do it q times, we would exhaust a. Recalling once again that $q = a \div b$, we have

> *$a \div b$ is the total number of groups when a objects are partitioned into equal groups of b objects*

This is called the *measurement* interpretation of division (in case there is no remainder). However, there is another common way in which we use division.

> **Activity:** In a third grade textbook, division is introduced as follows:
>
> > You can use counters to show two ways to think about dividing.
> >
> > (A) Suppose you have 18 counters and you want to make 6 equal groups. You can DIVIDE to find how many to put into each group.
> >
> > (B) Suppose you have 18 counters and you want to put them into equal groups, with 6 counters in each group.
>
> Although you can see easily in this special case that the answer to both problems is 6, discuss which of these two divisions uses the measurement interpretation and which requires a new understanding of division. It may help you to think more clearly if we replace 18 by 4023, and 6 by 27 in the above problem.

Suppose as usual we have a objects and $q = a \div b$. Suppose we divide the a objects into b equal groups, how many are in each group? Let $a = 15$ and $b = 5$ so that $3 = 15 \div 5$. We know that 15 can be partitioned into 3 groups of 5's, and we represent it pictorially by dots as follows:



Now if we count the dots by columns, we get 5 groups of 3's and $15 = 3 + 3 + 3 + 3 + 3 = 5 \times 3$. Thus if we divide 15 ($= a$) objects into 5 ($= b$) equal groups, there will be 3 objects in each group. But of course, 3 is just the quotient of $15 \div 5$, which by definition is equal to 3×5 $(= qb$. Comparing with $15 = 3 + 3 + 3 + 3 + 3 = 5 \times 3$, we immediately recognized that the commutativity of multiplication $3 \times 5 = 5 \times 3$ is at work.

In general then, with a objects and $q = a \div b$, then $a = qb$. By the commutativity of multiplication, $a = qb = bq = q + q + \cdots + q$ (b times), so that if we divide a objects into b equal groups, there will be q objects in each group. This leads to the following: assuming $q = a \div b$, then

> $a \div b$ *is also the number of objects in each group when* a *objects are divided into* b *equal groups.*

This is called the *partitive* interpretation of division (in case there is no remainder). *The fact that this interpretation is valid (i.e., yields the same number as the measurement interpretation) is due to the fact that multiplication among whole numbers is commutative.* For example, in the preceding Activity, task (A) requires that you use the partitive interpretation of division while task (b) requires the measurement interpretation.

At the risk of pointing out the obvious, both meanings of division are common place in everyday life. Suppose you give a party and you make a bowl of punch. If you want to find out how many cups of punch there are in the bowl, you are making a measurement division of the amount of fluid in your punch bowl by the amount of fluid in your cup. On the other hand, if four people decide to drink up the whole bowl of punch but wish to exercise caution by first computing how much fluid each must be prepared to take in if each drinks an equal amount, then these people will be doing partitive division of the amount of fluid in your punch bowl by 4. Everywhere you look, you will find both kinds of division done around you.

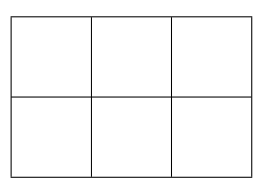
Example There is no better illustration of the two meanings of division than the problem of traveling (or motion). Suppose a car goes from town A to town B *at a constant speed*, which means that the distance traveled within *any* one-hour time interval is always a fixed constant. (Warning to the reader: while this may seem like a good example of contextual learning, we should not delude ourselves into believing that this is anywhere close to a "real world" situation. Drastic oversimplifications are involved. For instance, one rarely manages to drive at a constant speed for more than a few minutes in real life. There is also the implicit idealization in that the car is driving on a freeway that connects the two towns in a straight-line. How often does this happen in everyday driving?) A typical question is then the following:

suppose the distance between towns A and B is 264 miles, and the car gets to town B after 4 hours, what is the speed? If 264 miles is covered in 4 hours, we just partition 264 into 4 equal parts and the number of miles in one part is the number of miles traveled n one hour ("equal parts" because speed is assumed constant). *This is the partitive meaning of the division* $\frac{264}{4}$, which is 66. So the speed is 66 miles per hour. Another typical question is the following: suppose the speed is a constant 58 miles per hour and the distance between A and B is 522 miles, how many hours does it take to go from A to B? Here, we know that the car would be 58 miles from A after 1 hour, $58 + 58 = 116$ miles from A after 2 hours, $58 + 58 + 58 = 174$ miles from A after 3 hours, etc. So the question becomes how many 58's there are in 522. *This is the measurement interpretation of* $\frac{522}{58}$, which is 9. So it takes 9 hours to go from A to B.

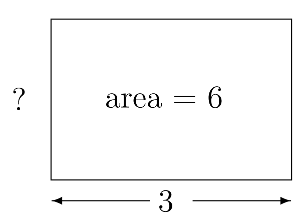
To summarize:

> *For motion in constant speed, computing the speed when distance and time are given is a partitive division problem, while computing the time to travel a certain distance at a given speed is a measurement division problem.*

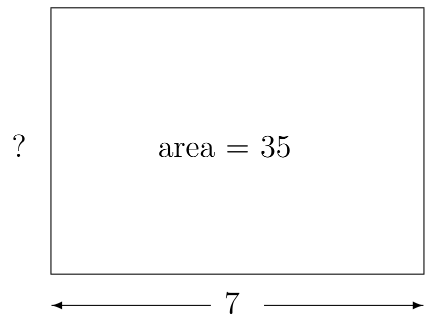
Finally, we give another geometric interpretation of division without remainder. In §2, we introduced an area model for multiplication. According to this model, 2×3, for example, would be modeled as the area of the rectangle with vertical side equal to 2 and horizontal side equal to 3:



Now suppose we ask for $6 \div 3 = ?$ From the point of view of the area model, this means we have a rectangle with area equal to 6 and a horizontal side equal to 3, and we want to know what the length of the vertical side is:



Similarly, the division $35 \div 7$ may be interpreted as asking for the length of the vertical side of a rectangle with area 35 and with the horizontal side equal to 7:



Of course, for whole numbers, such a geometric interpretation of division is no more than slightly entertaining. However, when we come to the division of fractions, the geometric interpretation would acquire added significance.

Exercise 3.20 Is 24 the quotient of $687 \div 27$? Is 13 the quotient of $944 \div 46$? Explain. (No calculator allowed.)

Exercise 3.21 Is 6977 the remainder of $124968752 \div 6843$? Why? (No calculator is allowed.)

Exercise 3.22 . *By taking multiples of the divisor*, find the quotient and remainder in each of the following cases: $964 \div 31$, $517 \div 19$, $6854 \div 731$, $4972086 \div 873$, and $4972086 \div 659437$.

Pedagogical Comment: The use of (only) a four-function calculator for the last two items involving 4972086 is allowed. However, it would be instructive to first ask for a ballpark figure of the quotient without the use of a calculator. This is a good exercise in making estimates, and would save a lot of guess-and-check in getting the correct quotient in each case. It should also be mentioned that there is an effective way to use the (four-function) calculator to get the quotient and remainder without any trial and error. How to do this and why it is true should lead to an interesting classroom discussion (one that in fact presupposes some knowledge of decimals). **End of Pedagogical Comment.**

Exercise 3.23 Let r be the remainder of $a \div b$. Suppose $a = mA$ and $b = mB$ for some whole numbers m, A and B. Let R be the remainder of $A \div B$. What is the relationship between R and r? Give a detailed explanation of your answer. (Caution: This problem is deceptive because it seems almost trivial, but the explanation is actually quite subtle and it requires the use of the uniquness of both the quotient and remainder which is discussed in the indented fine-print passage of this subsection.)

Exercise 3.24 You give your fifth grade class a problem:

> A faucet fills a bucket with water in 30 seconds, and the capacity of the bucket is 12 gallons. How long would it take the same faucet to fill a vat with a capacity of 66 gallons?

How would you *explain* to your class how to do this problem?

Exercise 3.25 Consider the following two problems: (a) If you try to put 234 gallons of liquid into 9 vats, with an equal amount in each vat, how much liquid is in each vat? (b) If you try to pour 234 gallons of liquid into buckets each with a capacity of 9 gallons, what is the minimum number of such buckets you need in order to hold these 234 gallons? Get the answer to both, and explain in each case whether you are using the partitive or the measurement interpretation of division.

3.5 The long division algorithm

Suppose we have to do the division problem $7864 \div 19$. Up to this point, the only way we can do it is to look at all the multiples of 19 until we get a whole number q so that $q \times 19 \leq 7864$ but $(q+1) \times 19 > 7864$. Of course we could painstakingly go through *all* the multiples one by one until we hit one with the above property, but that would be dull bookkeeping rather than mathematics. Let us do better. First of all, we can ignore small multiples like 10×19 or even 100×19 because all we care about is getting a multiple of 19 that is close to 7864. Let us make an estimate: the 100th multiple of 19 is 1900, so the 400th multiple is 7600, which is close to 7864. Add ten more of 19 and we get: $7600 + 190 = 7790$, which is even closer to 7864. A little experimentation shows that $7790 + 3 \times 19 = 7847 < 7864$ and $7790 + (19 \times 4) = 7866 > 7864$. Because $7790 + (3 \times 19) = (410 \times 19) + (3 \times 19) = 413 \times 19$ by the distributive law, we know 413 is the quotient. This method of finding the quotient is clearly superior to the monotonous checking of all the multiples of 19 and deserves to be made more systematic. With a little more work, this line of thinking would lead us to the long division algorithm, which is a beautiful and sophisticated method of finding the quotient and the remainder of a division-with-remainder.

Given two whole numbers a, b, with $b > 0$, our goal is to find an *efficient* algorithm that produces the quotient and remainder of $a \div b$. According to the second form of the division-with-remainder, this is the same as finding a q and an r so that

$$a = qb + r \qquad \text{and} \qquad 0 \le r < b$$

(See (35).) Let us illustrate the algorithm we are after, *long division algorithm*, by something relatively simple, say $a = 586$ and $b = 3$. Without further ado, here is the usual schematic presentation of this algorithm for $586 \div 3$:

$$\begin{array}{ccccc}
 & & 1 & 9 & 5 \\ \cline{2-5}
3 &) & 5 & 8 & 6 \\
 & & 3 & & \\ \cline{3-5}
 & & 2 & 8 & 6 \\
 & & 2 & 7 & \\ \cline{3-5}
 & & & 1 & 6 \\
 & & & 1 & 5 \\ \cline{3-5}
 & & & & 1
\end{array} \tag{37}$$

(Note that for reason of clarity of exposition, we bring the 6 down at each step of the long division in (37).) We know the conclusion we are supposed to draw from this: the mechanism described in (37) produces the quotient 195 and remainder 1. In other words,

$$586 = (\boxed{195} \times 3) + \boxed{1} \tag{38}$$

The question is *why*? Because this question can be easily misunderstood, let us explain it further. The question here is not why (38) is correct; the correctness of (38) is easily checked, after all, by verifying that $195 \times 3 = 585$ so that adding 1 to it produces 586. The question is rather why *the particular procedure adopted in (37), seemingly unrelated to multiplication or division in the usual sense we understand it*, should produce the correct answer of 195 and 1 in (38). The failure to directly address this question in school mathematics and pre-service professional development materials is what makes the long division algorithm so notorious. There is nothing wrong with the algorithm; it has already been stated above that it is one of the most beautiful pieces of elementary mathematics. There is plenty that is wrong with the way elementary mathematics is taught, however. Let us proceed to make amends by explaining the underlying reason why (37) leads to (38) simply and correctly.

First, we make a general comment about the long division algorithm n order to clarify our subsequent discussion. You can see from (37) that, in arriving at the purported quotient 195, the division-with-remainder is used three times: $5 \div 3$, $28 \div 3$, and $16 \div 3$. Now you may find the following fact puzzling: if we are trying to find an algorithm which is more efficient than the division-with-remainder itself in order to get at the quotient and the remainder, how can we be using the division-with-remainder itself? Here is the main point: for "small" numbers, the quotient and remainder of a division can be easily guessed at (e.g., $5 \div 3$, $28 \div 3$, and $16 \div 3$), so what the long division algorithm does is to break up the division of a large number (e.g., 586, although you can easily put up a number as large as you want) into a sequence of divisions of smaller numbers, and then string the latter together in an artful way so as to get at the quotient and remainder of the division of the original large number.[12] Our task is to understand why this "stringing together" makes sense.

Now we are going to give a *preliminary* explanation of why the steps in (37) lead to the correct conclusion (38). It is, we emphasize, only a preliminary explanation, because we shall subsequently point out in what way it is *unsatisfactory*.

The first step in (37) looks like $5 \div 3$, but since 5 stands for 500 in 586, this particular division is really $500 \div 3$, and the remainder 2 is really 200. So the first step in (33) is actually a restatement of the division-with-remainder $500 = (\boxed{100} \times 3) + \boxed{200}$. Therefore:

$$\begin{aligned} 586 = 500 + 86 &= \{([100 \times 3] + 200)\} + 86 \\ &= (100 \times 3) + 286\ . \end{aligned}$$

(At this point, we interject a word of caution: there will be many computations of this type in this subsection, and we have to resist the temptation

[12] It is a basic strategy in mathematics to try to break up a complicated task into a series of simpler tasks.

of multiplying out thing like 100×3 above, or 90×3 in the succeeding sentence. The reason is that we want the final outcome to be $(195 \times 3) + 1$ as in (38), and for this reason we want to keep the factor 3 intact all through these computations.) The second step in (37) is $28 \div 3$, which as before is in reality $280 \div 3$. The division-with-remainder in this case reads: $280 = (\boxed{90} \times 3) + \boxed{10}$. Thus,

$$\begin{aligned} 586 &= (100 \times 3) + 280 + 6 \\ &= (100 \times 3) + ([90 \times 3] + 10) + 6 \\ &= (100 \times 3) + (90 \times 3) + 16. \end{aligned}$$

The last step is easy: $16 = (\boxed{5} \times 3) + \boxed{1}$. So,

$$\begin{aligned} 586 &= (100 \times 3) + (90 \times 3) + (5 \times 3) + 1 \\ &= ([100 + 90 + 5] \times 3) + 1 \\ &= (\boxed{195} \times 3) + \boxed{1}\ , \end{aligned}$$

which is exactly (38). We have of course used the distributive law, and the reason why this law must play a role is not entirely obvious. It has to do with (35), where we have pointed out that multiplication and division undoes each other (at least in the case of no remainder). As we stressed in §3.3, the key reason underlying the multiplication algorithm is the distributive law. It therefore follows that *if division undoes multiplication, and the multiplicative algorithm depends critically on the distributive law, the long division algorithm must likewise make critical use of the distributive law.*

We next give an interpretation of (37) in terms of money. Suppose we have \$586 consisting of

5	hundred-dollar bills
8	ten-dollar bills
6	one-dollar bills.

Although we are trying to find out how many 3's there are in 586, we can turn the original problem around: suppose there are n 3's in 586 (with 0 or 1 or 2 left over), then we may also interpret n as the number of dollars in a stack when $3n$ dollars are divided equally into 3 stacks. To this end, we begin the process of creating these 3 stacks by first distributing the 5 hundred-dollar bills equally into these 3 stacks. In each stack we put in 1 hundred-dollar bill, and there are 2 left over. This corresponds to the first step of (37). Next, we convert the 2 hundred-dollar bills into 20 ten-dollar bills, so that (together with the original 8 ten-dollar bills already there) we now have 28 ten-dollar bills. These can be distributed into these three stacks equally with 9 in each stack and 1 left over. This corresponds to the second step of (37). Finally, we convert the 1 ten-dollar bill into 10 one-dollar bills, and we now have 16 one-dollar bills. Again, we can distribute them equally into the three stacks with 5 one-dollar bills in each, and 1 is left over. Altogether then, the original stack of $586 has been divided into three equal stacks each consisting of 1 hundred-dollar bill, 9 ten-dollar bills, and 5 one-dollar bills, with 1 one-dollar bill left over. This is exactly what (38) says.

Observe that the preceding interpretation of (37) in terms of money is very similar in spirit to the intuitive approach to finding a quotient described at the beginning of this sub-section. There are some people in mathematics education who consider the use of money to interpret the long divison algorithm as the height of conceptual understanding. Readers of this monograph cannot fail to realize, however, that the use of money is only an aid to understanding and is not to be confused with genuine mathematical understanding itself. We will in fact take up the **mathematical** explanation of long division next.

To make sure that the basic facts of the long division algorithm and the preliminary explanation are understood, we do another example: $1215 \div 35$:

$$\begin{array}{cccccccc} & & & & & 3 & 4 \\ \cline{4-8} 3 & 5 &) & 1 & 2 & 1 & 5 \\ & & & 1 & 0 & 5 & \\ \cline{4-8} & & & & 1 & 6 & 5 \\ & & & & 1 & 4 & 0 \\ \cline{4-8} & & & & & 2 & 5 \end{array} \tag{39}$$

Here is an abbreviated explanation:

$$\begin{aligned} 1215 &= 1210 + 5 \\ &= \{(30 \times 35) + 160\} + 5 \\ &= (30 \times 35) + 165 \\ &= (30 \times 35) + ([4 \times 35] + 25) \\ &= ((30 + 4) \times 35) + 25 \\ &= (\boxed{34} \times 35) + \boxed{25} \end{aligned}$$

For an in-depth understanding of the long division algorithm, the preceding analysis falls short in two respects. First, it lacks simplicity. It does not lead from a clear description of the algorithm straight to the desired conclusion about quotient and remainder; in fact, a clear description of the algorithm was never given. Second, the explanation does not clearly expose the role played by the sequence of remainders (i.e., the numbers 2, 1, 1 in (37) and the numbers 16 and 25 in (39)) which are critical to the understanding of the conversion of fractions to decimals in §4 of Chapter 4. We now give a mathematical explanation that is free of these defects.

Let us revisit (37). As we have emphasized throughout our discussion of algorithms, every one of the standard algorithms gains efficiency and simplicity by *ignoring place value* and by performing the operations one digit at a time, *mechanically*. The long division algorithm is the most remarkable embodiment of these features among the algorithms we have studied. Look at the "dividend" 586 (i.e., the number to be divided), and we shall describe the long division algorithm precisely in this special case. The idea will be seen to be prefectly general. We repeat: the algorithm will go through each digit of 586, one at a time, with absolutely no thought given to "breaking the dividend into parts" (as is taught in the schools).

There will be a sequence of steps, each of which performs a division-with-remainder **The divisor will always be the original divisor** (in this case it is 3). So the only thing that needs to be specified in each step is the dividend. We start from the *left*, for a change, and the dividend of the first step is the first digit of the original dividend (in this case it is the digit 5 of 586). More formally:

> *Step 1: perform the division-with-remainder, using as dividend the leftmost digit of the original dividend.*

So the first division is $5 \div 3$. The division-with-remainder gives

$$5 = (\boxed{1} \times 3) + \boxed{2}$$

The next (second) step is the crucial one, because the algorithm will be repeating this step ever after. The description of the dividend in the second step is this:

> *Step 2: Multiply the remainder of the preceding step by 10 and add to it the next digit (to the right) in the original dividend.*

In the present situation, the remainder from the first step is 2, and the next digit of the original dividend 586 is 8. So the number in question is $(2 \times 10) + 8 = 28$. Now divide 28 by the same divisor 3:

$$28 = (\boxed{9} \times 3) + \boxed{1}$$

We are now on automatic pilot:

Step 3: repeat step 2.[13]

With this in mind, the next digit in the original dividend 586 is 6, so the dividend of the next step is $1 \times 10 + 6 = 16$. Thus the third step of the long division algorithm is:

$$16 = (\boxed{5} \times 3) + \boxed{1}$$

Now (37) is entirely encoded in the following three (simple) division-with-remainders:

$$\begin{aligned} 5 &= (\boxed{1} \times 3) + \boxed{2} \\ 28 &= (\boxed{9} \times 3) + \boxed{1} \\ 16 &= (\boxed{5} \times 3) + \boxed{1} \end{aligned} \tag{40}$$

You could not possibly fail to observe that the quotient 195 is clearly displayed in (40) — read vertically down the first digits of the right sides — as well as the remainder 1 (the last term of the last equation). Though of slightly less interest, you can also read off the original dividend by going down the left sides of these equations and pick out the last digit of each number (in this case, you get 586). If you are baffled by these equations, let us hasten to point out that there is no mystery to them at all. You can, for example, relate them to the naive interpretation of (37) in terms of money in the following way. The first equation is a restatement of the splitting of the 5 hundred-dollar bills into three equal stacks with 2 left over. The second is the splitting of the 28 ten-dollar bills into three equal stacks with 1 left over, and the third is the splitting of the 16 one-dollar bills into three equal stacks with also 1 left over. But of course one must keep in mind that (40) is true regardless of any such monetary interpretations.

As we have emphasized, an algorithm is a sequence of mechanical procedures. Steps 1–3 explain how to generate these procedures as we go through the digits of the original dividend one-by-one. In the case of $586 \div 3$, (40)

[13] Recall what was said at the beginning of §3, to the effect that each standard algorithm would break a computation down to computations with single digit numbers.

gives the procedures explicitly. We now show how to *generate* (38) by the use of the long division algorithm as encoded in (40):

$$\begin{aligned} 586 &= (5\times 10^2) + (8\times 10) + 6 \\ &= \{([3\times 1] + 2)\times 10^2\} + (8\times 10) + 6 \end{aligned}$$

by the first equation of (40). This gives an explicit support to the interpretation about splitting the 5 hundred-dollar bills into three equal stacks with 2 left over. But to continue:

$$\begin{aligned} 586 &= (3\times 10^2) + (2\times 10^2) + (8\times 10) + 6 \\ &= (10^2\times 3) + (20\times 10) + (8\times 10) + 6 \\ &= (10^2\times 3) + (28\times 10) + 6 \\ &= (10^2\times 3) + ([3\times 9] + 1)\times 10 + 6 \end{aligned}$$

by the second equation of (37), The last line corresponds to the splitting of the 28 ten-dollar bills into tghree equal stacks with 1 left over. Now apply the last equation of (40) to get:

$$\begin{aligned} 586 &= (10^2\times 3) + ([9\times 10]\times 3) + 16 \\ &= (10^2\times 3) + ([9\times 10]\times 3) + (5\times 3 + 1) \\ &= (10^2\times 3) + ([9\times 10]\times 3) + (5\times 3) + 1 \\ &= \{(10^2 + [9\times 10] + 5)\times 3\} + 1 \\ &= (195\times 3) + 1\,. \end{aligned}$$

Let us review what we have accomplished. First we have verified (38) directly from a clearly stated digit-by-digit description of the long division algorithm, which is (40). Second, (40) exhibits the sequence of remainders 2, 1, 1 of the long division algorithm (in the case of $586\div 3$) which will be of critical importance later in Chapter 4. Third, there is a point which has been purposely suppressed thus far in order to get across the main thrust of the argument as clearly as possible, but which needs to be aired now. It is the fact that

> *at each step of the algorithm* ((37) or (40))*, the quotients 1, 9, and 5 are always single digit numbers.*

Obviously you have taken this for granted all along, because this is never mentioned in school mathematics. Nevertheless, the simplicity of the long division algorithm depends on "replacing" each digit of the original dividend by another digit (e.g., 5 by 1, 8 by 9 and 6 by 5; of course on other occasions, a digit of the dividend could be replaced by 0) so that this "single-digit" fact should be carefully verified. This we proceed to do. It is straightforward to see from (40) — and this is another reason why the explicit description of the algorithm in Steps 1–3 is critical — that because each remainder is by definition < 3, it is ≤ 2, so that the next dividend (i.e., the number 28 after the first step, and the number 16 after the second) is $\leq$ (20 + a single digit number) according to Step 2. Therefore the dividend at each step of the long division algorithm is always < 30. But if a number is < 30, then its quotient when divided by 3 must be a single digit, exactly as claimed. (Cf. the first form of the Division-with-Remainder in §3.4). As usual, we note that this reasoning is valid in general.

We want to drive home the point that the long division algorithm is strictly a digit-by-digit procedure without regard to place value. To this end we shall compare (37) with the division $58671 \div 3$, where the "dividend" 58671 has been chosen on purpose to have the three digits 5, 8, and 6 at the beginning. The usual schematic display of the long division is as follows:

$$
\begin{array}{rcccccc}
 & & 1 & 9 & 5 & 5 & 7 \\ \cline{2-7}
3 &) & 5 & 8 & 6 & 7 & 1 \\
 & & 3 & & & & \\ \cline{3-4}
 & & 2 & 8 & & & \\
 & & 2 & 7 & & & \\ \cline{3-5}
 & & & 1 & 6 & & \\
 & & & 1 & 5 & & \\ \cline{4-6}
 & & & & 1 & 7 & \\
 & & & & 1 & 5 & \\ \cline{5-7}
 & & & & & 2 & 1 \\
 & & & & & 2 & 1 \\ \cline{6-7}
 & & & & & & 0
\end{array}
$$

The precise description of the algorithm in accordance with Steps 1–3 is then:

$$\begin{array}{rcl} 5 & = & (\boxed{1} \times 3) + \boxed{2} \\ 28 & = & (\boxed{9} \times 3) + \boxed{1} \\ 16 & = & (\boxed{5} \times 3) + \boxed{1} \\ 17 & = & (\boxed{5} \times 3) + \boxed{2} \\ 21 & = & (\boxed{7} \times 3) + \boxed{0} \end{array} \qquad (41)$$

Now compare the first three steps of both (40) and (41). *They are identical.* For definiteness of discussion, let us concentrate on the third step in both. The 6 in 586 is in the ones digit, which is very different from the 6 in 58671, which is in the hundreds digit. In other words, the third step in (41) is actually $1600 \div 3$ and the corresponding division algorithm is then

$$1600 = (500 \times 3)100$$

if place value is taken into consideration. The point we wish to emphasize is that, *as far as the long division algorithm itself is concerned, place value is irrelevant*. Needless to say, the explanation of why (41) leads to $58671 = 3 \times 19557$ — along the line of the argument leading from (40) to (38) — is squarely based on (41) and nothing else. Contrast this with the first explanation given of (37), which is laden with place value interpretations. This is the point we made earlier about the lack of simplicity in the latter explanation.

As a reminder: observe once again that in (41), the dividend 58671 can be read off by going down the last digits of the left sides, and the quotient 19557 can be read off by going down the first digits of the right sides. Moreover, the remainder (0) appears in the last equation.

Let us give an example of a long division where the divisor has more than one digit, carried out in accordance with Steps 1–3, in order to illustrate more clearly why it is unnecessary to worry about "breaking the dividend into parts". Here then is $11546 \div 19$:

$$
\begin{array}{ccccccc}
 & & 0 & 0 & 6 & 0 & 7 \\ \cline{3-7}
1 & 9 & \big) 1 & 1 & 5 & 4 & 6 \\
 & & 0 & & & & \\ \cline{3-4}
 & & 1 & 1 & & & \\
 & & & 0 & & & \\ \cline{3-5}
 & & 1 & 1 & 5 & & \\
 & & 1 & 1 & 4 & & \\ \cline{3-6}
 & & & & 1 & 4 & \\
 & & & & & 0 & \\ \cline{5-7}
 & & & & 1 & 4 & 6 \\
 & & & & 1 & 3 & 3 \\ \cline{5-7}
 & & & & & 1 & 3
\end{array}
$$

Now the digit-by-digit sequence of steps are the following:

$$
\begin{aligned}
1 &= (\boxed{0} \times 19) + \boxed{1} \\
11 &= (\boxed{0} \times 19) + \boxed{11} \\
115 &= (\boxed{6} \times 19) + \boxed{1} \\
14 &= (\boxed{0} \times 19) + \boxed{14} \\
146 &= (\boxed{7} \times 19) + \boxed{13}
\end{aligned}
\tag{42}
$$

Now we use (42) to verify that $11546 = 19 \times 607 + 13$. Note that only the third and fifth equations carry real information, so only those will be used in the following derivation :

$$
\begin{aligned}
11546 &= 1 \times 10^4 + 1 \times 10^3 + 5 \times 10^2 + 4 \times 10 + 6 \\
&= (115 \times 10^2) + (4 \times 10) + 6 \\
&= (10^2 \times 115) + (10 \times 4) + 6 \\
&= (10^2 \times ([19 \times 6] + 1)) + (4 \times 10) + 6 \\
&= (6 \times 10^2) \times 19) + (10^2 + [4 \times 10]) + 6 \\
&= (6 \times 10^2) \times 19) + 146 \\
&= (6 \times 10^2) \times 19) + (7 \times 19) + 13 \\
&= (([6 \times 10^2] + 7) \times 19) + 13 \\
&= (607 \times 19) + 13
\end{aligned}
$$

Hence the quotient of $11546 \div 19$ is 607, and the remainder is 13.

Exercise 3.26 Explain to a sixth grader why the long division algorithm for $642 \div 4$ is correct.

Exercise 3.27 Compute $10192 \div 8$ using the long division algorithm. Then write out the *procedural description* of this long division along the line of (40), and use it to explain why your result is correct, i.e., if the quotient and remainder of your original long division are respectively q and r, use your procedural description to show directly that $10192 = (q \times 8) + r$.

Exercise 3.28 Do the same for $21850 \div 43$. Be sure you write down every step of the procedural description (as in (42), for example).

Exercise 3.29 Use (41) to derive the fact that $58671 = 3 \times 19557$. (In other words, we know you can multiply 3×19557 to get 19557, but we'd prefer that you learn how to use a sequence of division-with-remainders such as (41) to explain the long division algorithm.)

Exercise 3.30 Do the long division of $50009 \div 67$ to find the quotient and remainder, describe the algorithm as a sequence of division-with-remainders in accordance with Steps 1–3, and use these to show why your quotient and remainder are correct.

4 The Number Line and the Four Operations Revisited

We introduced the number line in §2 as an "infinite ruler" with the whole numbers identified with a set of equally spaced points (often referred to as "markers") to the right of a point designated as 0. A whole number n is also identified with the length of the line segment $[0, n]$ from 0 to n. Until the end of Chapter 4, we shall be concerned exclusively with the part of the number line to the right of 0.

In the following, we shall always refer to *the* number line, and this terminology has to be understood in the following sense. The positions of the whole numbers depend completely on the choice of 0 and 1. Once these two numbers have been fixed, the positions of the other whole numbers are likewise fixed. The segment $[0, 1]$ is called the *unit segment*, and the number 1 is sometimes referred to as *the unit*. In each discussion,

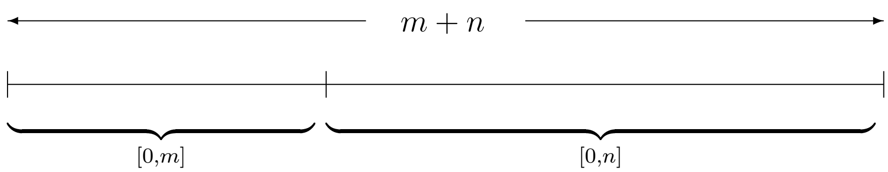
> *we always assume that a unit segment has been chosen on the given straight line so that the whole numbers are fixed on the line.*

It is in this sense that the number line is fixed. There will be occasions when we see fit to change the unit, in which case there will be a different number line to deal with. Such an occasion will come up soon enough.

The four arithmetic operations have also been interpreted geometrically. For addition, we have that for any two whole numbers m and n,

$$m + n = \text{the length of the segment obtained by concatenating the segments } [0, m] \text{ and } [0, n] \quad (43)$$

Geometrically, we have:

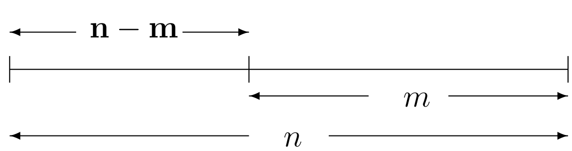


This way of adding numbers is exactly the principle underlying the slide rules of yesteryear. (For those who do not know what a "slide rule" is, perhaps one can describe it as the stone-age version of the calculator.) In any case, one can do many activities with (43) until this geometric way of adding numbers become second nature.

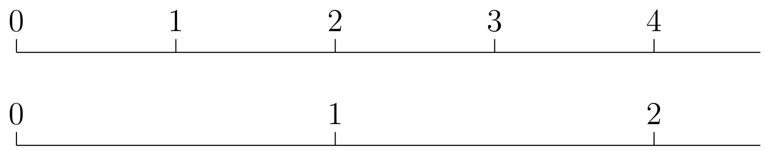
Subtraction is next. If m, n are whole numbers and $m < n$, then we saw in §3.2 that

$$n - m = \text{the length of the segment obtained when a segment of length } m \text{ is removed from one end of a segment of length } n \quad (44)$$

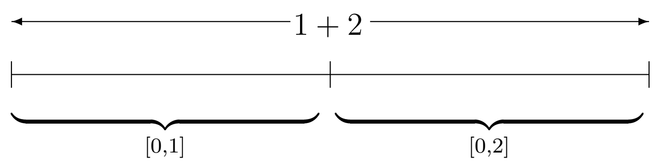
In picture:



We wish to go into some of the fine points of the addition and subtraction of whole numbers, especially with respect to the number line. Suppose we have two number lines, as indicated below.



What do you say to a student if she tells you that she gets $1 + 2 = 2$ in the following way: She takes $[0, 1]$ from the lower number line and $[0, 2]$ from the upper number line, and concatenate them as shown:



According to the lower number line, the resulting segment has length 2, and this is how she gets 2 for her answer. The question is: what is wrong?

In order to explain to her what the mistake is, you would have to recall for her the fact that all geometric representations of operations on whole numbers, including (43) and (44), are done *on one number line*, and therefore are done with respect to a *fixed* unit segment. So what she did wrong was not to realize that she had changed her unit segment in going from the upper number line to the lower number line, and subsequently got the two unit segments mixed up.

This brings up a fundamental issue in the arithmetic operations on whole numbers concerning

the importance of having the same unit as a fixed reference.

Consider for example the following equations:

$$\begin{aligned} 9 - 2 &= 1 \\ 8 + 16 &= 2 \\ 19 + 17 &= 3 \end{aligned} \qquad (45)$$

Although every equation in (45) is wrong according to the arithmetic of whole numbers as we know it, it is not as absurd as it appears. What (45) wants to say is the following:

$$\begin{aligned}
9 \text{ days} - 2 \text{ days} &= 1 \text{ week} \\
8 \text{ months} + 16 \text{ months} &= 2 \text{ years} \\
19 \text{ eggs} + 17 \text{ eggs} &= 3 \text{ dozen eggs}
\end{aligned}$$

The point of (45) is to underline the implicit or explicit role played by the unit in any addition or subtraction of numbers. The concept of a *whole number* is an abstract one: for example, the equation $2+3=5$ could mean any of the following among numerous other possibilities:

2 apples and 3 apples are the same as 5 apples
2 cups of coffee and 3 cups of coffee are the same as 5 cups of coffee
2 square inches and 3 square inches are the same as 5 square inches

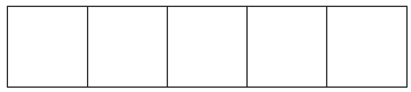
Whatever the interpretation of the abstract operations, each addition or subtraction must refer to the same unit. Thus, the first interpretation of $2+3=5$ is based on taking 1 to be "one apple", the second on "one cup of coffee", and the third on "one square inch". However, it can never be interpreted to mean

2 apples and 3 cups of coffee are the same as 5 square inches

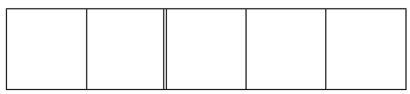
It is for the same reason — changing the unit in the middle of addition — that the student's reasoning of $1+2=2$ is wrong.

In a mathematical context, what we are saying is this. Suppose a ring $\mathbf{R}$ is isomorphic to the integers $\mathbb{Z}$, and suppose under the isomorphism $\bar{n} \in \mathbf{R}$ corresponds to $n \in \mathbb{Z}$. Then although both $\bar{2}+\bar{5}$ and $2+5$ make perfect sense, we cannot perform the addition $\bar{2}+5$ or $2+\bar{5}$.

This discussion of units assumes special importance when we discuss length and area. In this context, we come to understand (43) as the interpretation of 1 as a fixed unit length, so that each addition is nothing but combining unit lengths of segments. Now suppose we decide that 1 is the area of a fixed square. Recall the convention that once we decide on a fixed segment as the unit segment, then the area of the *unit square* (= the square with each side of length equal to that of the unit segment) defines the *unit area*. So the number "1" will henceforth refer to this unit area. Then the number 3 is no longer a concatenation of three unit segments but rather the total combined area of three unit squares. For convenience, we shall agree to interpret a whole number n in this context as the area of a rectangle whose width is a unit segment and whose length is the concatenation of n unit segments. For instance, 5 would be the area of the following rectangle:

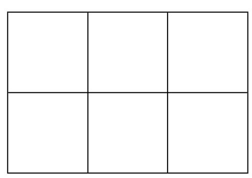


If there is no fear of confusion, we would simply say *this is a rectangle of width 1 and length 5.* Again, in this particular context, $2 + 3$ would be a concatenation of the following two rectangles rather than a concatenation of the segments $[0, 2]$ and $[0, 3]$ as stated in (43):



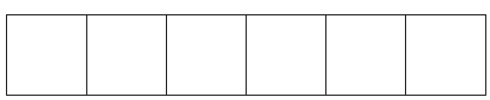
(46)

What makes this discussion particularly relevant is that we have interpreted the multiplication of whole numbers in §2 as area, e.g., 2×3 is the area of the rectangle:



(47)

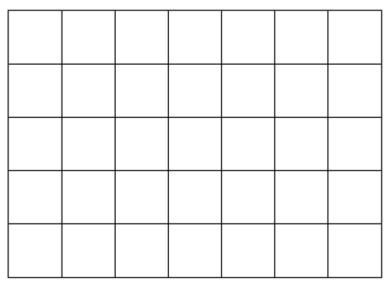
Therefore $2 \times 3 = 6$ means that the area of the the rectangle of (47) is the same as the area of



because the latter is exactly what 6 stands for in this context. (Or, to remind you of the meaning of the equal sign "=" as explained at the beginning of §2, the equality $2 \times 3 = 6$ means if we count the number of unit squares on the left and count the number of unit squares on the right, the two numbers are the same.) To drive home this point, we give one possible interpretation of $5+(2\times 3)=11$: it means that the area obtained by combining the rectangle of (46) and the rectangle of (47) is the same as the area of the rectangle whose width is 1 and whose length is 11. We shall be discussing the area interpretation of numbers extensively in Chapter 2.

Two extra comments would help to clarify the circle of ideas in connection with the representation of 1 as the area of the unit square.

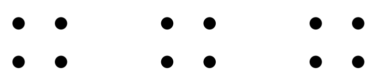
The first one is that declaring the area of the unit square (recall: this is the square whose side has length 1) to be 1 is nothing more than a *convention*. In fact, we could declare its area to be any number, say, 2. What could go wrong then? The area formula for the rectangle would be messed up, as follows. Look at the area of the rectangle with width 5 and length 7, for example. This rectangle is paved (tiled) by 5×7 unit squares:



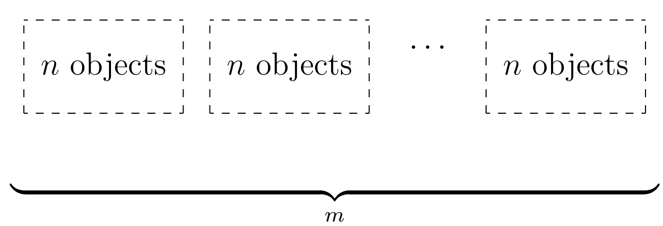
Therefore the area of the rectangle is $2 \times (5 \times 7)$, and not (5×7) (which would be the case if each unit square has area equal to 1). More generally, if a rectangle has width m and length n, then its area would be $2mn$ intead of the usual mn. Thus declaring the area of the unit square to be anything other than 1 would serve no purpose other than messing up otherwise simple formulas. For this reason, we all agree to set the area of the unit square to be 1.

A second comment is that the representation of the product of whole numbers as the area of the corresponding rectangle has a long history behind it. Until the time of Descartes (1596-1650), this was the *only* way to understand the multiplication of numbers. In the most influential mathematics textbook of all time, Euclid's *Elements* (circa 300 B.C.E.), there was never any mention of multiplying two numbers m and n. Each time Euclid wanted to express that idea, he would say: "the rectangle contained by the line m and the line n" (translation: "the rectangle" in Euclid means "the area of the rectangle", "contained by" means "having for its sides", and "the line m" means "the line segment of length m"). For this same reason, a product of three numbers, such as $12 \times 7 \times 9$, had to be interpreted as volume and therefore the product of four or more numbers was almost never considered until Descartes pointed out that multiplication can also be regarded as an abstract concept independent of geometry. Nowadays, a (good) college course on number systems would develop all number concepts in an abstract setting without reference to geometry. For this monograph, however, the geometric interpretation of multiplication not only is convenient for our purpose, but has the added advantage in that it is sufficiently similar to the common manipulative of Base Ten Blocks to make a beginner feel at ease. The purely algebraic approach to multiplication will be discussed in §7.2 of Chapter 2 and also §3 of Chapter 5.

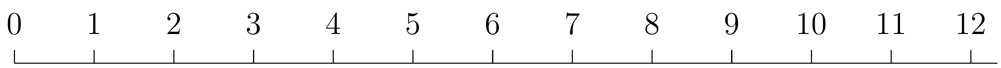
At this point, we are in a position to give another interpretation of the multiplication of whole numbers, one that was mentioned in §2. The product 3×4, for example, can be interpreted as the number 3 on a number line whose unit 1 is taken to be (the *magnitude* or *size* represented by) the number 4; one can think of 4 as a bag of 4 potatoes, a car full of 4 people, a box of 4 crayons, a bag of 4 marbles, etc. Schematically, in terms of such a choice of the unit 1, the number 3 is the point on the line represented by the following 3 groups of objects:



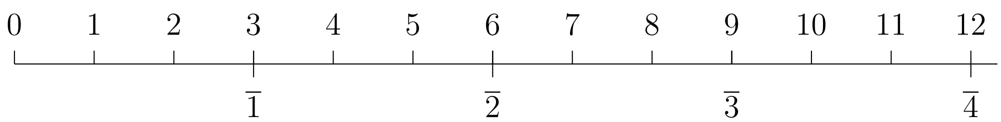
In general, if m and n are whole numbers, then mn may be interpreted as the number m on the number line whose unit 1 is taken to be (the *magnitude* or *size* represented by) the number n. So m in this context is the point on the line represented by the following m groups of objects:



For a later need, it would be advantageous to formalize this procedure. To facilitate the discussion, let us give m and n explicit values, say $m = 4$ and $n = 3$, and we shall re-interpret 4×3 as the point 4 on a number line with a new unit. So we start with a number line:



Now introduce new markers on the same line, where the new unit is 3. To avoid confusion, we shall distinguish the new number markings from the original one by a bar and place them underneath the line. So $\overline{1}$ is right under 3, $\overline{2}$ is right under 6, etc. In particular, 4×3 is four copies of the new unit and is therefore $\overline{4}$ in the new number line, and $3m$ would be just $\overline{m}$ for any whole number m.



This idea of using a *new unit* to re-interpret the multiplication of numbers provides an alternative way to understand the multiplication of fractions, as we shall see in §7.3 of Chapter 2.

Finally, division. For whole numbers a, b, with b always assumed to be nonzero, let us assume for now that a is a multiple of b, say $a = qb$ for some whole number q. Then

$$\text{we write} \qquad a \div b = q \qquad \text{for} \qquad a = qb. \tag{48}$$

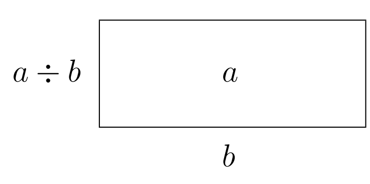
In intuitive language, multiplication and division undoes each other. (See (36) and the discussion surrounding it.) Sometimes, we also say multiplication and division are *inverse* operations to express this fact.

> **Activity:** Suppose a fourth grader understands $24 \div 3 = 8$ only in terms of the measurement interpretation and the partitive interpretation of division. Explain to her why $24 \div 3 = 8$ is the same as $24 = 8 \times 3$.

This is the place to tie up a loose end mentioned in the discussion of the division algorithm, namely, *why one cannot divide by 0*. Suppose, division by zero makes sense for a particular nonzero whole number n, say $n \div 0 = 3$. If one gives a little thought to what $m \div n = k$ could mean regardless of what m, n, k may be, one would likely conclude (with (48) as guide) that it means $m = nk$. In other words, one would require that (48) makes sense for all m, n, and k. Such being the case, $n \div 0 = 3$ would be the same as saying $n = 0 \times 3 = 0$, which contradicts our assumption that n is nonzero. If 3 is replaced by any other whole number, the same argument applies. So $n \div 0$ cannot be equal to any whole number if n is nonzero, which is another way of saying it cannot be defined. What about $0 \div 0$? We now run the preceding argument backwards, i.e., $m = nk$ should mean the same thing as $m \div n = k$ for all m, n. Thus knowing $0 = 0 \times 1$ means $0 \div 0 = 1$. But it is also true that $0 = 0 \times 2$, so $0 \div 0 = 2$. We have therefore shown that the value of $0 \div 0$ is ambiguous; it could be 1, or 2, or in fact any whole number by the same argument. This shows that $0 \div 0$ cannot be given a definite value, i.e., it is also undefinable. We have therefore shown that *division by 0 cannot be defined*.

It may be instructive to follow literally the partitive and measurement interpretations of division to see why $n \div 0$ cannot be defined for a nonzero whole number n. Suppose it were definable. By the partitive interpretation, $n \div 0$ would mean the number of objects in a part when n objects are partitioned into 0 equal parts (see §3.4). Because we cannot partition anything into 0 equal parts, this has no meaning. Now suppose $n \div 0$ were meaningful in the measurement sense. Then it is the number of parts when n objects are partitioned into different parts so that each part has exactly 0 objects. But if each part has no object, the partition of the n objects cannot be done. So again, this interpretation has no meaning either.

The fact that division undoes multiplication (always understood in the sense of (36) or (48)) leads to a geometric interpretation of division that has already been mentioned at the end of §3.4. Assuming as always that a is a multiple of b, then $a \div b$ is the other side of a rectangle whose area is a and one of whose sides is equal to b:



(49)

To conclude this discussion of division, it remains to point out that although the restriction that a is a multiple of b imposed here seems too severe, it will be seen when we come to §8 of Chapter 2 that it is in fact no restriction at all once fractions are at our disposal. The equivalence of division and multiplication as described in (48) will be seen to be the key to the understanding of division in general.

Exercise 4.1 Use the idea of introducing a new unit to represent a whole number c to re-interpret the distributive law in the form of $(a+b)c = ac+bc$.

Exercise 4.2 If a rectangle has area 98 and one side equals 14, what is the other side? If the area is 1431 and one side is 27? And if the area is 7797 and one side is 113?

5 What Is a Number?

Thus far, we have never paused to ask what a whole number is. We took this concept for granted from the beginning, and subsequently make them correspond to a collection of equally spaced markers (points) on a line to the right of a point denoted by 0. We are going to change our viewpoint here and use a set of markers on a line to *define* the whole numbers. Before you ask why we bother, let us do it first. So we start afresh by imposing a set of markers on a line.

> Take a straight line and mark off a point as 0 (zero). Then *fix a segment to the right of 0* and call it the *unit segment*. Mark the right endpoint of this segment on the line, thereby generating the first marker. Slide the unit segment to the right until its left endpoint is at the first marker; mark the new position of the right endpoint of the unit segment, thereby generating the second marker. Now slide the unit segment to the right again until its left endpoint rests on the second marker, and mark the right endpoint of the unit segment in its new position. This generates the third marker, etc. This generates a sequence of equally spaced markers to the right of 0.

Notice that up to this point, there is no mention of whole numbers. Now we will formalize the introduction of whole numbers by adopting the following definition.

Definition. A *whole number* is one of the markers on the line, so that, starting with the initial number 0, the next one (to the right of 0) is 1, the one after that is 2, etc., and we continue the naming of the markers in the same way we did the counting of the whole numbers in §1. This line with the whole numbers on it is called the *number line*. A *number* is by definition any point on the number line.

As far as whole numbers are concerned, what we have defined is at least consistent with everything we have done up to this point. In this sense, we are not by any means trying to attempt a retrograde revision of our knowledge of the whole numbers. Rather, we are saying that, for the work we do from now on, we shall agree to change our point of view and base our reasoning with whole numbers on this definition alone. Thus a whole number is now something very *concrete and explicit*: it is *among the markers on the number line* which were carefully constructed above. Because we have been using these markers all along in our work, no revision of anything we have done is necessary. Whatever we do in the future about whole numbers, however, we should be able to explain it in terms of these points on the number line.

You must be muttering to yourself at this point and wondering what is happening here. After all, don't we know what a whole number is? Let us define the number 5, for instance. Note that what is required here is not a description of our intuitive feelings about "5", but rather a precise definition analogous to the definition (say) of a *triangle* as three noncollinear points together with the line segments joining them. We know "five fingers". We also know "five chairs", or "five apples", or five of anything we see or touch because we can count. But *five* itself, without reference to any concrete object? So you see that it is difficult. Do not be discouraged, because the general concept of a number, in the sense defined above as a point on the number line, baffled the human race for over two thousand years before it was finally pinned down in the late nineteenth century. What we need for elemetary mathematics is fortunately nothing very sohpisticated, just the whole numbers and some other numbers which come out of whole numbers in a rather simple-minded fashion. Fractions or rational numbers, for instance. In other words, we will not scrutinize *every* point on the number line. Only a small portion of those points. This definition of numbers is not ideal, but it serves our pedagogical needs admirably, in the sense that it is accessible and it lends itself to a reasonable treatment of rational numbers and decimals. See Chapters 2, 4 and 5.

A precise definition of whole numbers is not strictly necessary if all we ever do in mathematics is to stay within the realm of whole numbers. For example, even if we cannot define precisely what 5 is, we can communicate the essence of it by putting up one hand with the fingers outstretched; that should be enough to communicate any kind of "fiveness" needed for conceptual understanding. This is the advantage of whole numbers: each has (at least in principle) a concrete manifestation such as outstretched fingers that almost renders abstract considerations about whole numbers unnecessary in elementary school. But we cannot stay with whole numbers forever, because the next topic is fractions. What concrete image can one conjure in connection with $\frac{13}{7}$ or $\frac{119}{872}$? Children need answers to this question in their quest for knowledge because they need something to anchor the many concepts related to fractions in the same way a handful of fingers can anchor any discussion about 5. Amazingly, school mathematics in our country has contrived to never answer this question. The results are entirely predictable: when adults abrogate their basic responsibilities, the first victims are the children. The generic non-learning of fractions among children has become part of our national folklore, so much so that you can find references to it in the comic strips of **Peanuts** and **FoxTrot**. We want to change this dismal scenario by adopting the down-to-earth approach of

giving direct answers to direct questions.

We will define fractions, decimals, or any concept we ever take up.

Now you would want to know why not just define fractions abstractly and leave a nice subject like whole numbers alone. The answer is that in order for children to understand fractions, fractions cannot be suddenly presented to them out of the blue. Learning is a gradual process firmly rooted in prior experiences. If we can convince them that fractions are nothing more than an natural extension of the whole numbers, then our chances of success in teaching them fractions would be immeasurably increased. At the moment, most (or perhaps all) of the school textbooks and professional development materials would have you believe that whole numbers are simple, but fractions are a completely different breed of animals. Whole numbers are taught one way, and fractions in a completely different way. There is no continuity from one to the other. The minute you as a teacher or your students buy into this view of numbers, our mathematics education is already in trouble because this means you have bought into **mathematical misinformation**. From the point view of mathematics, whole numbers are on an equal footing with fractions. They are part of the same family, the real numbers. In fact, this is where the number line comes in: we already have the whole numbers there, and the next step is to single out the fractions on this line. For this reason, it is not possible to only offer a precise definition of fractions and leave the whole numbers unattended. We must begin with a definition of the whole numbers that will naturally lead to fractions.

Let us consider the philosophical question of why something as natural as a whole number should be made into something as cold and formal as "a point on the number line". The answer lies in the fact that we are trying to deepen our understanding of the whole numbers in order to lay the groundwork for working with fractions. *In particular, the geometric interpretations* (43)–(44)*,* (46)–(47) *and* (49) *are part of this groundwork.* Now it is in the nature of human affairs that each time we try to achieve excellence in any endeavor, doing what is *natural* is simply not enough. Take running, for instance. This is about as natural an activity as we are going to get. In fact, had our ancestors been less good at it, all of them would have been hunted down by the predators on the African Savannas and we wouldn't be here to talk about fractions. Yet, if you talk to an Olympic sprinter, what you hear from him about running would strike you as extremely unnnatural if not downright unreal. He would tell you that he calculates exactly how many strides he takes with each breathe-in, how many he takes with each breathe-out, exactly how far from the starting block before he can take his first breath, and where the next spot is before he can take another breath again. Doesn't this unnnatural and calculated approach to running remind you of looking at a whole number as a point on the number line?

But let us not lose our perspective. Whatever the Olympic sprinters do in a race, it is highly unlikely that they think about "how soon before I can breathe in and how many strides I should take before then" each time they run to catch a bus. In the same way, you need not fixate on the number line every time you count oranges in the supermarket. All you need to do is to understand decimals and fractions and all the rest, and be good teachers. So please rise to the occasion when there is a need to regard a whole number as a point on the number line and be willing to work with this concept. If you can accept this reality, then you have already won half the battle.

6 Some Comments on Estimation (not yet written)

7 Numbers to Arbitrary Base (not yet written)