Taking Place Value Seriously: Arithmetic, Estimation, and Algebra

by Roger Howe, Yale University
and Susanna S. Epp, DePaul University

Introduction and Summary

Arithmetic, first of nonnegative integers, then of decimal and common fractions, and later of rational expressions and functions, is a central theme in school mathematics. This article attempts to point out ways to make the study of arithmetic more unified and more conceptual through a systematic emphasis on place-value structure in the base 10 number system. The article contains five sections.

Section one reviews the basic principles of base 10 notation and points out the sophisticated structure underlying its amazing efficiency. Careful definitions are given for the basic constituents out of which numbers written in base 10 notation are formed: *digit*, *power of 10*, and *single-place number*. The idea of *order of magnitude* is also discussed, to prepare the way for considering relative sizes of numbers.

Section two discusses how base 10 notation enables efficient algorithms for addition, subtraction, multiplication and division of nonnegative integers. The crucial point is that the main outlines of all these procedures are determined by the form of numbers written in base 10 notation together with the Rules of Arithmetic [1]. Division plays an especially important role in the discussion because both of the two main ways to interpret numbers written in base 10 notation are connected with division. One way is to think of the base 10 expansion as the result of successive divisions-with-remainder by 10; the other way involves successive approximation by single-place numbers and provides the basis for the usual long-division algorithm. We observe that long division is a capstone arithmetic skill because it requires fluency with multiplication, subtraction, and estimation.

Section three discusses how base 10 notation can be extended from nonnegative integers to decimal fractions – that is, fractions whose denominators are powers of 10. The procedures for the arithmetic operations also extend seamlessly. This is a practical reflection of the fact that the Law of Exponents can be extended to hold for all integers, not just nonnegative integers. The ease of computation when numbers are written in base 10 notation together with the efficient approximation properties discussed in sections 2D and 5, make decimal fractions an efficient tool for practical computation, especially since numbers coming from measurements are known only approximately. The key features of base 10 notation have led to its heavy exploitation in machine computation.

Section four explores the connections of base 10 notation with algebra. It is suggested that decimal numbers can be profitably thought of as "polynomials in 10," and parallels between base 10 computation and computation with polynomials are illustrated. Understanding the basic structural aspects of base 10 arithmetic can promote comfort with algebraic manipulation.

Section five discusses ordering, estimation, and approximation of numbers. For comparing numbers, the concept of *relative* place value is a crucial idea. The corresponding idea for approximation is *relative error*. In many cases, both relative and absolute error can be controlled by using base 10 expansions. A key concept is *significant figure*, and we note that relative accuracy of approximation improves rapidly – and relative error decreases rapidly – with the number of significant figures. In fact, it is usually unreasonable to expect to know a "real-world" number (meaning the result of a measurement) to more than three or four significant figures, and often much less accuracy is enough. Failure to appreciate the limits of accuracy may be one of

[1] By the "Rules of Arithmetic," we mean what mathematicians refer to as the Field Axioms, and what are often called "number properties," or just "properties." They are nine in number: four (Commutative, Associative, Identity and Inverse Rules) for addition, four parallel ones for multiplication, and the Distributive Rule to connect addition and multiplication. Other rules, such as "Invert and multiply," or the formula for adding fractions, or the rules of signs for dealing with negative numbers, can be deduced from these nine basic rules. See the Appendix for a more complete description of the Rules of Arithmetic.

the most pervasive forms of innumeracy: it affects many people who are for the most part quite comfortable with numbers. Scientific notation, which focuses attention on the size of numbers and the accuracy to which they are known, is also discussed, as is the question of accuracy and estimation in arithmetic computation.

We hope that the issues addressed in this article will be of interest to a broad spectrum of mathematics educators, a term we use inclusively, to comprise mathematics teachers at all levels, as well as others with a professional interest in mathematics education. Substantial effort has been devoted to making the contents accessible to a fairly wide audience, but what one reader may find obvious, another may find obscure, and vice versa. It is hoped that many readers can appreciate the broad message, that place value can serve as an organizing and unifying principle across the span of the elementary mathematics curriculum and beyond, and we beg our readers' indulgence with parts that may seem either over- or under- elaborated, keeping in mind that other readers may see them in the opposite light.

Acknowledgements: The authors are grateful for comments on earlier drafts of this article from the Mathematics School Study Group committee members, from Johnny Lott and his associates in NCTM and ASSM, especially Gail Englert, Bonny Hagelberger, and Mari Muri, and from Scott Baldridge, Thomas Roby, and Kristin Ulmann. Thanks to Mel Delvecchio and W. Barker for vital production help.

1. The Base 10 System: Single-place Numbers, Expanded Form, and Order of Magnitude

Our ordinary base 10 system is a highly sophisticated method for writing numbers efficiently. It uses only 10 symbols (the ***digits***: 0,1,2,3,4,5,6,7,8,9), arranged in carefully structured groups, to express any number. Further, it does so with impressive economy. To express the total human population of the world would require only a ten-digit number, needing just a few seconds to write.

The efficiency of the base 10 system is possible because of its systematic use of mathematical structure. We hope that making more aspects of this structure explicit will increase conceptual understanding and improve computational flexibility, thereby helping to make mathematics instruction more effective. It may also promote numeracy by making students more sensitive to order of magnitude, and to the estimation capabilities of place-value, or base 10, notation. Finally, it can illuminate the parallels between arithmetic and algebra, thus making arithmetic a preparation for algebra, rather than the impediment that it sometimes appears to be [KSF, Ch.8].

All the numbers discussed in this section and the next will be nonnegative integers. When such a number is written in base 10 notation, we will say that it is in ***base 10 form***. A number in base 10 form is implicitly broken up into a sum of numbers of a special type. For example, consider

$$7,452 = 7,000 + 400 + 50 + 2.$$

The right-hand side of this equation is usually called the ***expanded form*** of the number, and we often say that the digit 7 is in the ***thousands place***, 4 is in the ***hundreds place***, 5 is in the ***tens place***, and 2 is in the ***ones place***. Although the concept of expanded form is mentioned in state mathematics standards, this article explores what would follow from making it central to arithmetic instruction.

Each of the addends in the expanded form of a number will be called a ***single-place number***. Thus each single-place number can be written as a digit – possibly 0 – times a ***power of 10*** (that is, a product of a certain number of 10's).[2] The following display shows the various representations for the single-place numbers in the previous example:

$$\begin{array}{lllllll} 7,000 & = & 7 \times 1000 & = & 7 \times (10 \times 10 \times 10) & = & 7 \times 10^3 \\ 400 & = & 4 \times 100 & = & 4 \times (10 \times 10) & = & 4 \times 10^2 \\ 50 & = & 5 \times 10 & = & 5 \times 10^1 & & \\ 2 & = & 2 \times 1 & & 2 \times 10^0. & & \end{array}$$

[2]Exponential notation is discussed in section 2C.

The ***order of magnitude of a (non-zero) single-place number*** is the number of zeroes used to write it, or, equivalently, it is the exponent of 10 when the number is written using exponential notation. For example, since $7,000 = 7 \times 10^3$ and since $400 = 4 \times 10^2$, the order of magnitude of 7,000 is 3 and the order of magnitude of 400 is 2.

When a positive integer is written in base 10 form, its ***leading component*** is its largest single-place component, and its ***order of magnitude*** is the order of magnitude of its leading component. For instance, because the leading component of 7,452 is 7,000, the order of magnitude of 7,452 is 3. When the meaning is clear, we will sometimes refer to a number's "magnitude" rather than to its "order of magnitude."

A key point about expanded form is that no two single-place components of an integer have the same order of magnitude. In fact, this property characterizes expanded form: A sum of single-place components is the expanded form of a number exactly when

a. it involves exactly one component of each order of magnitude up to the magnitude of the number, and

b. the components are arranged from left to right in descending order according to their orders of magnitude.

It follows that the expanded form of an integer is unique.

Given the terminology we have introduced, we can say either that in the number 7,452 the digit 4 is in the hundreds place or that the single-place component with magnitude 2 in 7,452 is 400. Note, however, that although 7,052 does not have a single-place component with magnitude 2, we say that it has a zero in the hundreds place.

2. Single-place Numbers and the Algorithms of Arithmetic

The fact that numbers in base 10 form are sums has a pervasive influence on the methods for computing with them. Indeed, the usual procedures for adding, subtracting, multiplying, and dividing two such numbers are obtained by applying the Rules of Arithmetic to the sums of their single-place components. The Rules of Arithmetic play the critical role of reducing an arithmetic computation involving two numbers to a collection of simple single-digit computations involving the single-place components of the numbers.

2A. Addition

We call the basic strategy for adding two numbers in base 10 form the ***addition algorithm.*** It consists of three steps:

1. Break each of the numbers (the "addends") into its single-place components.

2. Add the corresponding components for each order of magnitude. If the component for an order of magnitude of an addend is missing, it is treated as if it were zero.

3. Recombine the sums from step 2 into a number in base 10 form. (The details of this step give rise to the trickier parts of the algorithm.)

Here is a simple example:

$294 + 603$	$= (200 + 90 + 4) + (600 + 3)$	by breaking each addend into its single-place components
	$= (200 + 600) + 90) + (4 + 3)$	by grouping the single-place components by order of magnitude
	$= 800 + 90 + 7$	by adding components for each order of magnitude
	$= 897$	by recombining the single-place components into a base 10 number.

To be able to analyze the steps of the addition algorithm, we need to develop some preliminary ideas.

Basic Rules for Addition: The basic strategy is justified by means of the two most basic rules of addition, the Commutative Rule and the Associative Rule, together with the Distributive Rule, which is the rule that involves both addition and multiplication.

The Commutative Rule for Addition: The value of a sum does not depend on the order of the addends. In other words, for any two nonnegative integers, a and b,

$$a + b \;=\; b + a.$$

The Associative Rule for Addition: When adding three numbers, the value of the sum does not depend on the way the numbers are combined into pairwise sums. More precisely, for any three nonnegative integers a, b, and c,

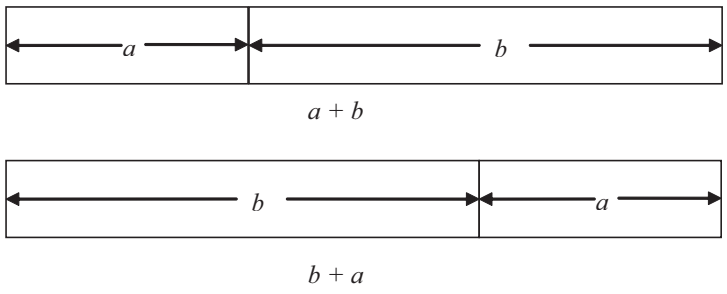
$$(a + b) + c \;=\; a + (b + c).$$

***The Distributive Rule*:** The product of one nonnegative integer times a sum of two others is obtained by adding the first number times the second plus the first number times the third. More precisely, for any nonnegative integers a, b, and c,

$$a \times (b + c) \;=\; a \times b + a \times c.$$

A variety of methods are commonly used to justify these rules. For instance, if we think of a number as corresponding to a length, say the length of a bar, then the process of adding two numbers is modeled by placing the corresponding bars end to end and measuring the total length. When this model is used, it is easy to see why the commutative and associative rules are true, as is illustrated in Figures 1 and 2.

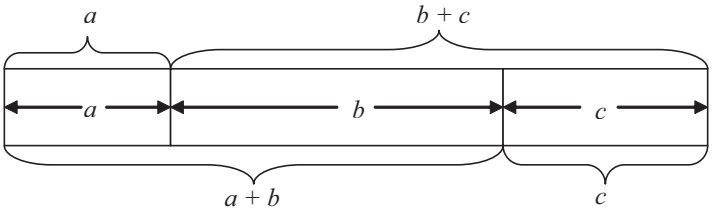
Figure 1: The Commutative Rule for Addition



The bar at the top of Figure 1 represents the sum $a + b$. If it is flipped across its midpoint, the order of the addends is reversed to become $b + a$, and yet the length remains the same.

Associativity is even simpler to justify: One just takes a bar composed of sub-bars of lengths a, b, and c, and observes that it can be thought of as being made of sub-bars of length $a + b$ and c, or, equally easily, of sub-bars of lengths a and $b + c$.

Figure 2: The Associative Rule for Addition

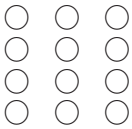


$(a + b) + c = a + (b + c)$

To illustrate the Distributive Rule, we often use an ***Array Model*** for multiplication. If a nonnegative integer is represented as a horizontal row of some sort of object, say small circles, then multiplication involves replicating the row a certain number of times. For instance, if 3 is pictured as

then 4×3 can be represented as 4 groups of 3:

For the Array Model, instead of arranging all the groups along the same line, we stack them one above the other, to make a rectangular array:

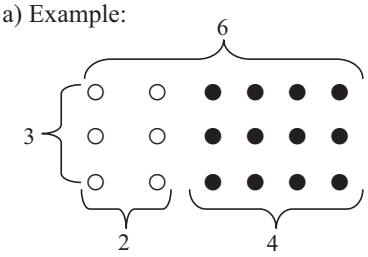


This is the Array Model for 4×3: four horizontal rows of three items each, stacked vertically above each other.

The Distributive Rule can be visualized in terms of the Array Model, as shown in Figure 3.

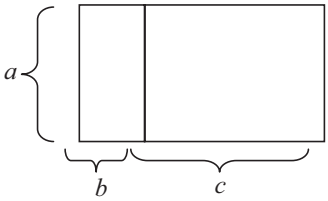
Figure 3: Distributive Rule

a) Example:



$3 \times (2 + 4) = 3 \times 2 + 3 \times 4$

b) Schematic Representation:



$a \times (b + c) = a \times b + a \times c$

The Commutative and Associative Rules for Addition are basic principles of arithmetic, but they are rarely used in isolation. Most practical calculations call for combinations of both of them. We express the result in a rule, whose name emphasizes the freedom that follows from using the two rules together.

***The Any-Which-Way Rule*:** When a list of numbers is to be added, it does not matter in what sequence we do the pairwise sums or how we order the addends in any of the intermediate sums: the final result will always be the same.

Although formal demonstrations of the Any-Which-Way Rule and the Distributive Rule are not appropriate when multidigit addition is first introduced, nevertheless explicit discussion about them, using examples, is not only feasible but advisable – particularly when the examples show how using them can simplify calculations and aid mental math.

Adding Two Numbers: The Simplest Case: How do the Any-Which-Way Rule and the Distributive Rule justify the strategy of adding two numbers in base 10 form? To answer this question, first look at the sum of two single-place numbers with the same order of magnitude. For example,

$$\begin{aligned} 7,000 + 2,000 &= 7 \times 1000 + 2 \times 1000 && \text{by rewriting 7,000 and 2,000 as products} \\ &= (7 + 2) \times 1000 && \text{by the Distributive Rule} \\ &= 9 \times 1000 && \text{by adding the digits 7 and 2} \\ &= 9,000. && \text{by rewriting } 9 \times 1000 \text{ in base 10 form.} \end{aligned}$$

These steps justify the ordinary procedure for adding two single-place numbers with the same order of magnitude: we simply add the digits and append the number of zeroes corresponding to the numbers' common order of magnitude. Thus, for instance, when adding 7 thousands to 2 thousands, the result is (7 + 2) thousands, or 9 thousands.

This observation can be combined with the Any-Which-Way Rule to add many pairs of numbers. For example,

$$\begin{aligned} 7,452 + 1,326 &= (7,000 + 400 + 50 + 2) + (1,000 + 300 + 20 + 6) \\ &\qquad \text{by converting to expanded form} \\ &= (7,000 + 1,000) + (400 + 300) + (50 + 20) + (2 + 6) \\ &\qquad \text{by the Any-Which-Way Rule to group together single-place components} \\ &\qquad \text{with the same order of magnitude} \\ &= 8000 + 700 + 70 + 8 \\ &\qquad \text{by the procedure for adding single-place numbers with the same order of magnitude} \\ &= 8,778 \quad \text{by converting to base 10 form.} \end{aligned}$$

.

Adding Two Numbers with Regrouping: The simple procedure shown in the preceding example does not always work because it is possible to add two single-place numbers of one order of magnitude and obtain a number with a larger order of magnitude. For example,

$$80,000 + 60,000 = 14 \times 10,000 = 140,000 = 100,000 + 40,000.$$

The result of this addition has two single-place components: one has the same order of magnitude as the addends and the other has order of magnitude one larger.

We can represent the general procedure for adding single-place numbers of the same order of magnitude as follows:

$$(a \times 10^n) + (b \times 10^n) = (a + b) \times 10^n,$$

where a and b are digits and n is a nonnegative integer. Because a and b are each less than 10, their sum is less than 20, but if their sum is 10 or more, then $(a + b) \times 10^n$ has order of magnitude greater than n. To be specific: When $a + b \geq 10$, then $a + b - 10$ is a number from 0 through 9, and

$$(a \times 10^n) + (b \times 10^n) = 10 \cdot 10^n + (a + b - 10) \times 10^n = 10^{n+1} + (a + b - 10) \times 10^n.$$

This result leads to the familiar need to "regroup" or "rename" or "carry." For example,

$$
\begin{aligned}
7,453 + 1,729 &= (7000 + 400 + 50 + 3) + (1000 + 700 + 20 + 9) \\
&\qquad \text{by converting to expanded form} \\
&= (7000 + 1000) + (400 + 700) + (50 + 20) + (3 + 9) \\
&\qquad \text{by the Any-Which-Way Rule to group together single-place components} \\
&\qquad \text{with the same order of magnitude} \\
&= 8,000 + 1,100 + 70 + 12 \\
&\qquad \text{by the procedure for adding single-place numbers with the same order of magnitude} \\
&= 8,000 + (1,000 + 100) + 70 + (10 + 2) \\
&\qquad \text{by writing 1,100 and 12 in expanded form} \\
&= (8,000 + 1,000) + 100 + (70 + 10) + 2 \\
&\qquad \text{by the Any-Which-Way Rule to group together single-place components} \\
&\qquad \text{of the same magnitude} \\
&= 9,000 + 100 + 80 + 2 \\
&\qquad \text{by the procedure for adding single-place numbers with the same order of magnitude} \\
&= 9,182 \qquad \text{by converting to base 10 form.}
\end{aligned}
$$

When two numbers are added, the sum of the components for any given order of magnitude contributes directly to this order of magnitude and, possibly, to the next larger one. Moreover, because the sum of two digits is at most 18 (the result of adding 9 plus 9), if adding two digits leads to a contribution to the next larger order of magnitude, it adds 1 to the sum of the digits of the components. For instance, in the example above adding the 3 and the 9 led to a result of 2 in the component with order of magnitude zero (the ones place) and an addition of 1 to the digits in the component with order of magnitude one (the tens place). Similarly, adding the 4 and the 7 in the components with order of magnitude two led to a result of 1 in the component with order of magnitude two (the hundreds place) and an addition of 1 to the digits in the component with order of magnitude three (the thousands place). Note that if there is a carry of 1 from the addition of two digits, the maximum sum of the digits for the next order of magnitude, including the carry of 1, is 19 (the result of adding 9 plus 9 plus 1). Thus, in all cases when two numbers are added, the maximum possible carry from one order of magnitude to the next is 1.

In general, therefore, an efficient way to find any sum of two numbers is to start with the components with order of magnitude zero (the ones place), find their sum, combine the 1 (if it occurs) with the digits of the components with order of magnitude one (the tens place), find the sum of those components, and so forth, progressing systematically, one by one, to larger orders of magnitude until the components of greatest order of magnitude in the numbers have been added.

Considerations for Teaching: Observe that the procedure described above is exactly what one does when executing the standard addition algorithm, although the role played by the expanded form of the numbers may be hidden when an overly procedural instructional approach is used. If manipulatives, such as Cuisenaire rods, are used to help make these ideas clear, it is essential that teachers take care to link their use to the way additions are represented on paper. How many students appreciate that when they put one number under the other with the ones digits aligned, the effect is to put the digits corresponding to the same order of magnitude under each other, so that adding the digits in the columns exactly accomplishes combining the components with the same order of magnitude? If the role of order of magnitude is not highlighted, students may come to believe that the procedure for adding decimal fractions (i.e., aligning the decimal points) is different from the procedure for adding nonnegative integers (i.e., right justifying the digits). On the other hand, if the procedures are understood in terms of combining components with equal order of magnitude, they will see that the two are exactly the same. (See section 4 for more discussion of decimal fractions.) Finally, note that the logical basis for the procedure to add many numbers by "adding the digits in the columns" is a natural extension of the procedure to add two numbers.

Thinking explicitly about the component-by-component aspect of base 10 addition can enable greater flexibility in doing addition. For example, when adding two-digit numbers mentally, it is usually easier to think of adding the 10s and the 1s separately (thinking of the numbers in expanded form). Then, one usually

adds the 10s first, since this gives the main part of the sum, and finally adds on the sum of the 1s. If there is a carry from the 1s, it just increases the result from the 10s addition by one more 10, which is easy to do when there are only two decimal places to worry about.

We will see later on (section 5) that these same considerations lead to quick estimation procedures for sums, easy enough to be carried out mentally in many cases.

The ideas about addition in this section need not be introduced in a single large lump. They can be taught as the base 10 system is now, in a gradual way, starting with two-digit numbers and observing that a two-digit number consists of a certain number of 10s and a certain number of 1s. (Explicit attention to this idea may be helpful because there is evidence that the irregularities of English two-digit number words, such as eleven and twelve, somewhat impede children's ability to think in term of place value [KSF, p.167].) The strategy of "adding the tens digits and adding the ones digits" should then appear more or less natural, at least after discussion and work with examples. Problems that do not require regrouping can be done first to establish the basic principle. Then the issue of what to do when one or the other of the digit sums is greater than 10 can be discussed.

Even before general two-digit addition is considered, it may be helpful to prepare for it by using the "make a ten" idea [KSF, p.189] for sums of two digits that add up to numbers greater than 10. As explained in [Ma], Chinese teachers think of the addition facts not simply as a list to be memorized, but as a way to begin introducing the structure of the base 10 system. When a sum of two digits is greater than 10, they encourage students to think of gathering together the parts that add up to 10 and imagining the sum as made up of that 10 together with the sum of the parts that are left over. For instance, they would think of $8+6$ as using 2 from the 6 to combine with the 8, making 10 with 4 left over. So the answer would be "four and ten," or "fourteen." In Chinese, this approach may have arisen naturally because all the words for the numbers from 11 through 19 express the number as 10 plus a digit. In English, the words for the numbers 13 through 19 can fairly easily be seen to split in this way, though not quite as obviously as in Chinese. But the connection is more obscure for the words eleven and twelve although these also evolved from words meaning one-over-ten and two-over-ten in Old English. Perhaps our students could get a better start in mathematics if we emphasized developing addition facts using the "make a ten" idea and helped them hear the way the number names reflect base 10 structure.

2B: Subtraction

Thinking in terms of single-place components can also help with subtraction.

Connecting Subtraction with Addition: Subtraction should be linked with addition at every opportunity so that students come to understand that $a-b=c$ *because* $b+c=a$. In other words, $a-b$ is what must be added to b to get back a. This approach prepares students to understand, when negative numbers are eventually introduced, that subtraction is the same as adding the additive inverse.

Linking subtraction with addition also has practical value. For one thing, it provides extra practice for learning families of addition facts so that, for example, $6+9=15$, $9+6=15$, $15-9=6$, and $15-6=9$ are seen as belonging to the same family. Moreover, the computational details in the problems $a-b=c$ and $b+c=a$ correspond closely: the subtraction $a-b=c$ involves regrouping (by borrowing) exactly when the addition $b+c=a$ involves regrouping (by carrying). Indeed, the same orders of magnitude are affected in both computations.

For example, in the problems

$$\begin{array}{r} 970,957 \\ \underline{+\ 50,829} \\ 1,021,786 \end{array} \qquad\qquad \begin{array}{r} 1,021,786 \\ \underline{-\ \ 50,829} \\ 970,957 \end{array}$$

the addition involves carrying from the 1s to the 10s, and from the 100s to the 1000s. Then a carry from the 10,000s to the 100,000s causes a further carry to the 1,000,000s. Correspondingly, in the subtraction, one has to borrow from the 10s to the 1s and from the 1,000s to the 100s, and one has to "borrow across a zero," from the 1,000,000s, to the 100,000s, to the 10,000s.

Borrowing Across a Zero and Rollover: The previous example includes one of the most troublesome aspects of subtraction, namely "borrowing across a zero." Study of this phenomenon in the context of the parallel addition problem reveals an interesting feature of addition, the "rollover phenomenon," that often goes unremarked. To help students understand it, it can be helpful to have them consider the differences in the two sets of computations shown below:

$$\begin{array}{rr} 658 & 999 \\ +341 & -341 \\ \hline 999 & 658 \end{array} \qquad (a) \qquad\qquad \begin{array}{rr} 658 & 1{,}000 \\ +342 & -342 \\ \hline 1{,}000 & 658 \end{array} \qquad (b)$$

In example (a), the digits in each place add up to 9, but there are no carries. In example (b), the digits in the tens and hundreds places add up to 9 but the digits in the ones places add up to 10. This produces a 0 in the ones place of the answer and a carry of 1 to the tens place of the sum. As a result, the digits in the tens places add up to 10, which produces a 0 in the tens place of the answer and a carry of 1 to the hundreds place of the sum. This makes the digits in the hundreds place add up to 10, which produces a 0 in the hundreds place and a 1 in the thousands place of the answer. We call this phenomenon ***rollover***. Its extreme form occurs in adding 1 to a number like 99 or 999 or 9999, where all the digits equal 9. The result, of course, is the denomination that is one larger than the given number. For instance:

$$\begin{array}{r} 999{,}999{,}999 \\ + \qquad\qquad 1 \\ \hline 1{,}000{,}000{,}000 \end{array}$$

In the days when car odometers were mechanical and an increase of one mile changed the odometer from 999 to 1000, or, from 9,999 to 10,000, or, especially, from 99,999 to 100,000, it was amusing to watch the 9's change one-by-one into 0's going from right to left across the display.

Borrowing across a zero is the parallel for subtraction to rollover for addition: when an addition $a+b=c$ involves rollover, the corresponding subtraction $c-b=a$ requires borrowing across a zero, and vice versa. If rollover is explicitly studied as an interesting and exceptional case that can occur in doing addition, and if subtraction is consistently connected to addition, then borrowing across a zero may seem less mysterious to students. Here is an example of rollover in a more typical addition, along with the corresponding borrowing across zeroes in the associated subtraction:

$$\begin{array}{rr} 3{,}537 & 6{,}001 \\ +2{,}464 & -2{,}464 \\ \hline 6{,}001 & 3{,}537 \end{array}$$

In the addition, the carry from 1s to 10s causes two places to roll over: it causes the 10s to carry to the 100s, and that makes the 100s carry to the 1,000s. In the subtraction, in order to subtract in the 1s place, one must borrow from the 1,000s place, across both the 100s and 10s places.

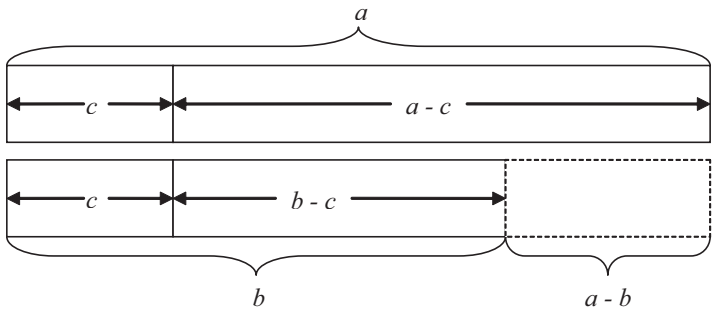
Alternative Methods for Subtraction: The previous discussion focused on the standard algorithm for subtraction, but thinking in terms of single-place components suggests two alternative algorithms plus a third that is a combination of the two. As is well-known (see [KSF] for example) there are multiple interpretations for subtraction. Two of these, the comparison and missing-addend interpretations, suggest additional ways to perform subtractions.

Compensation Method: Comparison involves a process of matching elements in two sets, followed by counting the unmatched elements of the larger set after all elements of the smaller set have been used up. If one thinks in this way, one can see that if the same amount is subtracted from two numbers, the difference of the results is the same as the difference of the original numbers. In algebraic terms, this property is expressed by the identity

$$a-b=(a-c)-(b-c),$$

which illustrates the fact that on the number line, if one shifts two numbers by equal amounts, the difference between them stays the same. This is shown in Figure 4.

Figure 4: Compensation Principle for Subtraction



This compensation principle for subtraction follows algebraically from the fact that $-(b-c) = -b+c$, from which it follows that the c's cancel out in the identity:

$$(a-c)-(b-c) = a-c-b+c = a-b.$$

The computational principle based on this identity is sometimes called a *compensation method*. It can be applied in various ways to simplify subtraction problems. For example,[3]

$$\begin{array}{rcr} 10,000 & \overset{\text{subtract 1}}{\rightarrow} & 9,999 \\ \underline{-\quad 964} & \overset{\text{subtract 1}}{\rightarrow} & \underline{-\quad 963} \\ & & 9,036 \end{array}$$

The point here is that subtracting 1 obviates the need for regrouping so that in the second problem, no borrowing is needed.

***Missing-Addend Method (Making Change)*:** Another way to solve a subtraction problem of the form $a-b$ is to figure out what would have to be added to b to obtain a. For example, consider

$$\begin{array}{r} 10,000 \\ \underline{-\quad 964} \\ 9,036 \end{array}$$

To find the answer, one can "count up," successively producing numbers divisible by higher and higher powers of 10. In this case, starting from 964, one adds 6 to get 970, then 30 to make $1,000$, and then (noting that one does not need to add any 100s to make a thousand) one adds $9,000$ to make the goal of $10,000$:

$$964 \; \{+ \; 6 \; (\rightarrow 970) + 30 \; (\rightarrow 1000) + 9,000 \; (\rightarrow 10,000)\}.$$

Thus, the missing addend is $6+30+9,000 = 9,036$, and so $10,000 - 964 = 9,036$.

A similar procedure is sometimes used by cashiers in making change. As cashiers hand back each piece of change to customers, they sometimes repeat out loud the sequence of numbers to be added to the purchase amount to add up to the size of the bill given in payment. The method amounts to doing the subtraction through a place-by-place calculation of the missing addends. However, because U.S. monetary denominations are 1 cent, 5 cents, 10 cents, 25 cents, 50 cents, 1 dollar, 5 dollars, and so forth, the making-change procedure

[3]Thanks to Mari Muri for bringing this to our attention. It has also been taught for many years by Herb Gross, among others.

is applied to subtractions from single-place numbers as well as from powers of ten. For example, making change for a purchase of \$2.85 from a \$5 payment, amounts to computing \$5.00 − \$2.85. One would start from \$2.85, then add 5 cents to reach \$2.90, 10 cents to reach \$3.00, and finally \$2.00 to reach \$5.00:

$$\$2.85\ \{+\ \$0.05\ (\rightarrow \$2.90) + \$0.10\ (\rightarrow \$3.00) + \$2.00\ (\rightarrow \$5.00)\}.$$

So the change is \$5.00 − \$2.85 = \$0.05 + \$0.10 + \$2.00 = \$2.15. In giving you the change, the cashier might start by saying, "Two-eighty-five," and then say, in handing you first the nickel, then the dime, and finally the two dollars, "Two-eighty-five, two-ninety, three dollars, five dollars."

***A Combination Method*:** Another subtraction procedure systematically utilizes place-by-place compensation to reduce a problem to a simpler one, in which, for each magnitude, the digit of either the subtrahend or the minuend is zero. The result is that the initial problem is decomposed into a set of missing-addend problems.

To be specific:

> Given a subtraction $c - a$ of numbers in base 10 form, for each order of magnitude, subtract from both numbers the smaller of the base 10 components of c and a with that magnitude. The result is an equivalent subtraction problem in which, for each order of magnitude, only one of the addends has a non-zero component.

For example, consider the problem of computing 83 − 26. The two components in the ones place (i.e., with magnitude zero) are 3 and 6, so 3 is subtracted from both 83 and 26 to obtain the equivalent problem 80 − 23. The components in the tens place (i.e., of magnitude one) are 80 and 20, so 20 is subtracted from both 80 and 23 to obtain the equivalent problem 60 − 03, which is computed using the missing-addend method, i.e., by answering the question "What must be added to 3 to obtain 60?":

$$\begin{array}{rcrcr} 83 & \overset{\text{subtract 3}}{\rightarrow} & 80 & \overset{\text{subtract 20}}{\rightarrow} & 60 \\ \underline{-\ 26} & \overset{\text{subtract 3}}{\rightarrow} & \underline{-\ 23} & \overset{\text{subtract 20}}{\rightarrow} & \underline{-\ 03} \\ & & & & 57 \end{array}$$

Now let's look at a three-digit problem where borrowing across a zero would ordinarily be needed. To compute 204 − 9, we note that the components in the ones place are 4 and 9, so we subtract 4 from both 204 and 9 to obtain the equivalent problem 200 − 5, which we may compute using the missing-addend method. Note how easily this procedure adapts to mental arithmetic.

$$\begin{array}{rcr} 204 & \overset{\text{subtract 4}}{\rightarrow} & 200 \\ \underline{-\ \ 9} & \overset{\text{subtract 4}}{\rightarrow} & \underline{-\ \ 5} \\ & & 195 \end{array}$$

When a problem is simplified using this version of the compensation method, the result can be computed as a sum of missing-addend problems. For example, consider the following five-digit subtraction:

$$\begin{array}{rcrcrcrcrl} 20{,}413 & \overset{\text{subtract 3}}{\rightarrow} & 20{,}410 & \overset{\text{subtract 400}}{\rightarrow} & 20{,}010 & \rightarrow & 20{,}000 & + & 10 & \\ \underline{-\ \ 906} & \overset{\text{subtract 3}}{\rightarrow} & \underline{-\ \ 903} & \overset{\text{subtract 400}}{\rightarrow} & \underline{-\ \ 503} & \rightarrow & \underline{-\ \ 500} & + & \underline{-\ 3} & \\ & & & & & & 19{,}500 & + & 7 & = 19{,}507 \end{array}$$

After the subtrahend and minuend were fully reduced, the answer was obtained by breaking the problem into the sum of 20000 − 500 and 10 − 3, each of which was computed using the missing-addend method. Note that another way to finish the computation would have been to break the problem into blocks and insert the separate answers into the original at the appropriate places:

$$20050 - 503 = (20000 - 500) + (10 - 3) = (200 - 5) \times 100 + 7 = 195 \times 100 + 7 = 19500 + 7 = 19{,}507.$$

2C: Multiplication

Thinking in terms of single-place numbers also sheds light on multiplication and division as well as on addition and subtraction. And just as the commutative and associative rules for addition are an important part of the procedures for addition, the commutative and associative rules for multiplication play a crucial role in the procedures for multiplication.

The Commutative Rule for Multiplication: The commutative rules for multiplication says that the order of the factors does not matter in computing a product of two numbers.

> ***The Commutative Rule for Multiplication***: For any two nonnegative integers n and m, the products $m \times n$ and $n \times m$ are the same:
> $$m \times n = n \times n.$$

Probably the best way to develop an intuitive sense for the rule is through the Array Model. This model is also important because it helps prepare students develop intuition for the concept of area. For example, consider the arrays used to represent 4×3 and 3×4:

Reflecting the diagram on the left across its diagonal produces essentially the same figure as the diagram on the right: 4 rows of 3 objects give the same total number of objects as 3 rows of 4 objects: $4 \times 3 = 3 \times 4$. The same reasoning can be applied to any array with n rows and m columns to help visualize the general Commutative Rule for multiplication.

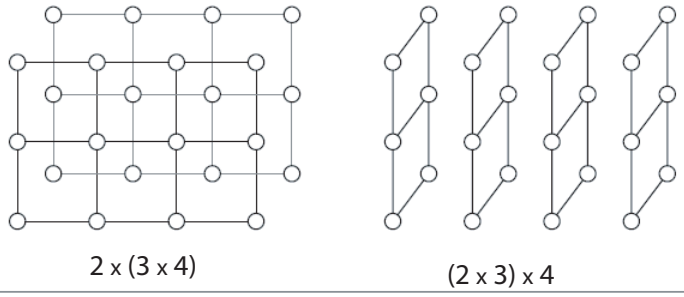
Associative Rule for Multiplication: The Associative Rule for Multiplication concerns the formation of products of three numbers. This is done in two stages: The product of two adjacent numbers is computed and then the result is multiplied by the third number. One can start either with the leftmost two numbers or with the rightmost two. The Associative Rule for Multiplication says whichever way one starts, one will always obtain the same result.

For example, suppose we want to multiply 2, 3, and 4 together. We can multiply the 2 and the 3 together, to obtain 2×3, and multiply the result by 4, or we can multiply the 3 and the 4 together, to obtain 3×4, and multiply the result by 2. Parentheses indicate the order, giving $2 \times (3 \times 4)$ for the first method and $(2 \times 3) \times 4$ for the second. The general statement of the property is as follows:

> ***The Associative Rule for Multiplication***: For any nonnegative integers ℓ, m, and n,
> $$\ell \times (m \times n) = (\ell \times m) \times n.$$

One can extend the Array Model to imagine $\ell \times (m \times n)$ as a rectangular 3-dimensional array consisting of ℓ layers of flat arrays of size $m \times n$. This 3-dimensional array can also be viewed as n perpendicular slices of flat arrays of size $\ell \times m$. This idea is illustrated in Figure 5 for $\ell = 2$, $m = 3$, and $n = 4$.

Figure 5: The Associative Rule for Multiplication



The Associative Rule for Multiplication is less obvious than the Associative Rule for Addition. Indeed, writing both sides of the property for almost any triple of numbers produces an equation that takes some thought to verify. For example, the product of $n = 5$, $m = 6$ and $\ell = 7$ can be computed as $(5\times6)\times7 = 30\times7$ or as $5 \times (6 \times 7) = 5 \times 42$. Thus, the Associative Rule asserts that $30 \times 7 = 5 \times 42$. While this equation is certainly true, if we did not know where it came from, we probably would have to compute both sides to check it.

Note that the Associative Rule for Multiplication may be applied successively to products of more than three factors. For instance,

			by the associative rule with
$2 \times (3 \times (4 \times 5))$	$=$	$2 \times ((3 \times 4) \times 5)$	$\ell = 3$, $m = 4$, and $n = 5$
	$=$	$(2 \times (3 \times 4)) \times 5)$	$\ell = 2$, $m = 3 \times 4$, and $n = 5$
	$=$	$((2 \times 3) \times 4) \times 5$	$\ell = 2$, $m = 3$, and $n = 4$
	$=$	$(2 \times 3) \times (4 \times 5)$	$\ell = 2 \times 3$, $m = 3$, and $n = 4$.

Any-Which-Way Rule for Multiplication: Once students have developed facility with the commutative and associative rules for multiplication, they can start working with the version of the Any-Which-Way Rule for multiplication:

> ***The Any-Which-Way Rule for Multiplication***: Not only can the factors in a product of several factors be grouped according to any scheme of parentheses, but also pairs of them can be interchanged in any way one desires.

For example,

$(3 \times 8) \times (7 \times 8)$	$=$	$3 \times (8 \times (7 \times 8))$	by the Associative Rule for Multiplication
	$=$	$3 \times ((8 \times 7) \times 8)$	by the Associative Rule for Multiplication
	$=$	$3 \times ((7 \times 8) \times 8)$	by the Commutative Rule for Multiplication
	$=$	$3 \times (7 \times (8 \times 8))$	by the Associative Rule for Multiplication
	$=$	$(3 \times 7) \times (8 \times 8)$	by the Associative Rule for Multiplication

Thus, in this example, we could just go ahead and interchange the 8 on the left with the 7 on the right. This is a special case of the following useful identity: For all numbers a, b, c, and d,

$$(a \times b) \times (c \times d) = (a \times c) \times (b \times d).$$

Someone who thinks that the transformations shown above amount to a lot of work to justify the simple exchange of factors would be right. Note, however, that when the first and last factors are multiplied out, the result is

$$24 \times 56 = (3 \times 8) \times (7 \times 8) = (3 \times 7) \times (8 \times 8) = 21 \times 64,$$

and the equality of the leftmost and rightmost products is not immediately obvious. In addition, the Any-Which-Way Rule is useful for mental arithmetic. For instance, a person who finds it difficult to compute $(7 \times 5) \times (6 \times 2)$ might see that it is much easier to find the product if the 5 and the 2 are grouped together to give a factor of 10:

$$(7 \times 5) \times (6 \times 2) = (7 \times 6) \times (2 \times 5) = 42 \times 10 = 420.$$

Exponential Notation: Given any positive integers b and n, we define b^n to be the number obtained by multiplying n copies of b together:

$$b^n = \underbrace{b \times b \times \cdots \times b}_{n \text{ factors}}$$

We also define b^0 to equal 1:

$$b^0 = 1.$$

In the expression b^n, b is called the ***base*** and n the ***exponent***. By the Any-Which-Way Rule for Multiplication, all the different ways we could group the b's as we compute b^n would give the same result. Another result of the Any-Which-Way Rule for Multiplication is the Law of Exponents.

> ***The Law of Exponents***: For any positive integer b and all nonnegative integers m and n,
> $$b^m \times b^n = b^{m+n}.$$

For example,

$$10^3 \times 10^5 = 10^{3+5} = 10^8, \quad 2^7 \times 2^4 = 2^{7+4} = 2^{11}, \quad 63^{25} \times 63^{52} = 63^{25+52} = 63^{77}.$$

Besides its utility in computation, the Law of Exponents reveals a beautiful parallel between multiplication and addition. The Rules of Arithmetic emphasize the fact that both addition and multiplication are governed by essentially the same four basic rules. This suggests that there might be some deep relationship between addition and multiplication. The Law of Exponents is a concrete expression of this relationship.

The Extended Distributive Rule: Recall from section 2A that the Distributive Rule tells us how to multiply a number times a sum of two numbers:

> ***The Distributive Rule*:** For any nonnegative integers a, b and c, we have the equation
> $$a \times (b + c) \;=\; a \times b \;+\; a \times c.$$

Just as repeated uses of the commutative and associative rules can be combined to produce an any-which-way rule, so repeated uses of the Distributive Rule, combined with the commutative and associative rules, produce the *Extended Distributive Rule*. To express this rule, we use a standard convention for indicating a product of two numbers: instead of writing a multiplication symbol between them we simply put the numbers side by side.

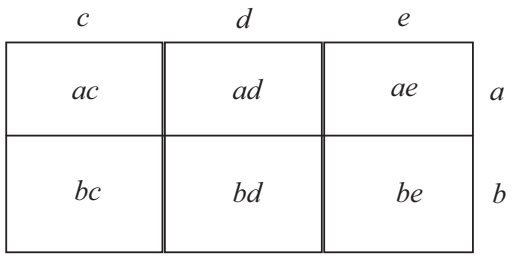
> ***The Extended Distributive Rule*:** If A and B are sums of several numbers, then the product AB may be computed by multiplying each addend of B by each addend of A, and adding all the resulting products.

For example, if $A = a + b$ and $B = c + d + e$, then

$$\begin{aligned} AB &= (a+b)(c+d+e) && \text{by substitution} \\ &= ac + ad + ae && \text{first } a \text{ is multiplied by } c, d, \text{ and } e \text{ successively} \\ &\quad + bc + bd + be. && \text{and then } b \text{ is multiplied by } c, d, \text{ and } e \text{ successively} \\ & && \text{and the results are added together.} \end{aligned}$$

A schematic version of the Extended Distributive Rule for multiplying a sum of 2 numbers by a sum of 3 numbers is shown in Figure 6.

Figure 6: Extended Distributive Rule (Schematic Form)



$$(a + b)(c + d + e) = ac + ad + ae + bc + bd + be$$

Applying the Rules to Compute Products: The Any-Which-Way rules for addition and multiplication plus the Extended Distributive Rule, which links multiplication and addition, are the tools we use for multiplying general numbers in base 10 form. Since each such number is the sum of its single-place components, the Extended Distributive Rule tells us that the product of two of them may be found by multiplying each single-place component of one factor by each single-place component of the other factor, and then adding all the products together. Moreover, the Any-Which-Way Rule for addition says that we have great latitude how we sum these numbers up. Differences in multiplication algorithms come mainly from choosing different summation procedures.

Multiplying Single-Place Numbers: As indicated by the previous paragraph, in order to compute products of numbers in base 10 form, it is essential to understand products of single-place numbers. The situation is somewhat analogous to that of addition:

> ***Products of Single-Place Numbers***: A product of two single-place numbers equals the product of the digits times the denomination whose magnitude is equal to the sum of the orders of magnitude of the factors. In other words, if $M = d_1 10^m$ and $N = d_2 10^n$, then
>
> $$MN = d_1 10^m \times d_2 10^n = d_1 d_2 \times 10^{m+n}.$$

For example,

$$\begin{aligned} 200 \times 4000 &= (2 \times 100) \times (4 \times 1000) \\ &\quad \text{by writing 200 and 4000 in expanded form} \\ &= (2 \times 4) \times (100 \times 1000) \\ &\quad \text{by the Any-Which-Way Rule for Multiplication}^4 \\ &= 8 \times 100,000 \\ &\quad \text{by computing 2 times 4 and 100 times 1000} \\ &= 800,000 \\ &\quad \text{by converting from expanded form.} \end{aligned}$$

Another way to express the computation uses exponential notation and the Law of Exponents:

$$\begin{aligned} 200 \times 4000 &= (2 \times 10^2) \times (4 \times 10^3) \\ &\quad \text{by writing 200 and 4000 in expanded form with exponents} \\ &= (2 \times 4) \times (10^2 \times 10^3) \\ &\quad \text{by the Any-Which-Way Rule for Multiplication} \\ &= 8 \times 10^5 \\ &\quad \text{by computing 2 times 4 and using the Law of Exponents} \\ &= 800,000 \\ &\quad \text{by converting from expanded form.} \end{aligned}$$

Observe that when two single-place numbers are multiplied, if the product of the digits is less than 10, the result is another single-place number whose magnitude is the sum of the magnitudes of the factors:

$$30 \times 200 \;=\; 6,000 \;=\; 6 \times 1000.$$

If the product of the digits is a multiple of 10, the result is a single-place number whose magnitude is one more than the sum of the magnitudes of the factors:

$$60 \times 500 \;=\; 30 \times 1000 \;=\; 30,000 \;=\; 3 \times 10,000.$$

And if the product of the digits is greater than 10 but not divisible by 10, a sum of two single-place numbers is obtained, the larger of which has magnitude that is one more than the sum of the magnitudes of the factors:

$$30 \times 700 \;=\; 21,000 \;=\; 20,000 + 1,000 \;=\; 2 \times 10,000 + 1,000.$$

Doing Multidigit Multiplication: We can now combine the observations about multiplying single-place numbers with the Extended Distributive Rule to do multidigit multiplication. As an example, consider 36×457. We first write each of these numbers in expanded form. If we then use the diagram for the Extended Distributive Rule, we have the array shown in Figure 7.

Figure 7: Extended Distributive Rule for 36 x 457

	400	50	7	
30	30 x 400	30 x 50	30 x 7	30
6	6 x 400	6 x 50	6 x 7	6

(30 + 6) x (400 + 50 + 7) = 30 x 400 + 30 x 50 + 30 x 7 + 6 x 400 + 6 x 50 + 6 x 7

Notice that the sum of the products in the top row of the array is the result of multiplying each single-place component of 457 by 30 and adding the results:

$$30 \times 400 + 30 \times 50 + 30 \times 7 \;=\; 12000 + 1500 + 210 \;=\; 13,710.$$

The sum of the products in the bottom row of the array is the result of multiplying each single-place component of 457 by 6 and adding the results:

$$6 \times 400 + 6 \times 50 + 6 \times 7 \;=\; (2,400 + 300 + 42) \;=\; 2,742.$$

These are exactly the numbers obtained for the partial products in a version of the standard algorithm for computing 36×457:

$$\begin{array}{r} 457 \\ \times\ \ 36 \\ \hline 2,742 \\ 13,710 \\ \hline 16,452 \end{array}$$

Similarly, if we interchange the order of the factors 36 and 457, we obtain

$$\begin{array}{r} 36 \\ \times 457 \\ \hline 252 \\ 1,800 \\ 14,400 \\ \hline 16,452 \end{array}$$

where the partial products in the third, fourth, and fifth rows are the sums of the products in the first, second, and third columns of the array shown in Figure 7. The sum of the products in the first column is the result of multiplying each single-place component of 36 by 400 and adding the results, the sum of the products in the second column is the result of multiplying each single-place component of 36 by 50 and adding the result, and the sum of the products in the third column is the result of multiplying each single-place component of 36 by 7 and adding the results:

$$\begin{array}{lcll} 30 \times 400 + 6 \times 400 & = & 400 \times 6 + 400 \times 30 & = & 14,400 \\ 30 \times 50 + 6 \times 50 & = & 50 \times 6 + 50 \times 30 & = & 1,800 \\ 30 \times 7 + 6 \times 7 & = & 7 \times 6 + 7 \times 30 & = & 252 \end{array}$$

Another way to think of this is that each of the partial products results from multiplying one of the single-place components of 457 times the number 36: $7 \times 36 = 252$, $50 \times 36 = 1,800$, and $400 \times 36 = 14,400$.

Interpreting the standard multiplication algorithm in terms of the array of products of single-place components could help students to understand why it doesn't matter which factor in a product comes first, even though the intermediate steps in the calculation are so different.

Alternative Algorithms for Multidigit Multiplication

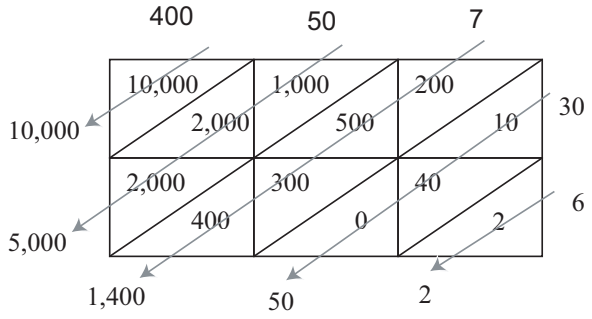
The Lattice Method: Given our base 10 notation, we have tried to show that the most natural way to add two numbers is to break each addend into its single-place components, sum up the components of each magnitude, and then recombine the results. Neither of the standard algorithms for multiplication, however, uses this approach. However, the idea does form the basis for an multiplication algorithm known as the Lattice Method (or the Array Method, or Napier's Bones). It is a refinement of the observation that if we write the products of the components in an array, as shown in Figure 8, then the magnitude of the products decreases from left to right along rows and from top to bottom along columns.

Figure 8: Row and Column Totals for 36 x 457

			Row Totals
30 x 400 = 12,000	30 x 50 = 1,500	30 x 7 = 210	13,710
6 x 400 = 2,400	6 x 50 = 300	6 x 7 = 42	2,742
Column Totals: 14,400	1,800	252	16,452 Grand Total

Observe in Figure 8 that the numbers that have the same magnitude appear along the diagonals of the array, going from upper right to lower left. We can refine this further by breaking up the product in each box into its single-place components, dividing each box by a diagonal line, and placing the tens digit of the digit product times the appropriate magnitude above the diagonal line and the ones digit times the appropriate denomination below the diagonal line. The result is that single-place components with the same magnitude are arranged along the diagonals, and the diagonals are composed of right triangles, facing alternately up and down, as shown in Figure 9. Summing along these diagonals allows us to add numbers that all have the same magnitude, and the sum of these sums gives the overall product.

Figure 9: The Lattice Model for 36 x 457



Grand Total: 10,000 + 5,000 + 1,400 + 50 + 2

We add up the digits along the diagonals, going from left to right, to obtain the following:

$$\begin{aligned} 36 \times 457 &= 10,000 + (2+2+1) \times 1,000 + (4+3+5+2) \times 100 + (0+4+1) \times 10 + 2 \\ &= 10,000 + 5 \times 1,000 + 14 \times 100 + 5 \times 10 + 2 \\ &= 10,000 + 5,000 + 1,400 + 50 + 2 \\ &= 16,452. \end{aligned}$$

Note that the process of summing along the diagonals in the Lattice Method is parallel to shifting the partial products to the left in the standard algorithm.

The Every-Product Method: Another multiplication algorithm, now used in some elementary school textbooks, could be called the Every-Product method. It has children write all the individual single-place components separately before adding them together. This parallels the way we usually do addition, namely by breaking each addend into its single-place components, adding all the components of each magnitude, and combining the results. To begin, let's show one way to obtain all the single-place-component products:

$$\begin{aligned} 36 \times 457 &= (30+6) \times (400+50+7) \\ &\qquad \text{by definition of base 10 notation} \\ &= 30 \times 400 + 30 \times 50 + 30 \times 7 + 6 \times 400 + 6 \times 50 + 6 \times 7 \\ &\qquad \text{by the Extended Distributive Rule} \\ &= 12000 + 1500 + 210 + 2400 + 300 + 42 \\ &\qquad \text{by the rules for multiplying single-place numbers.} \end{aligned}$$

The ***Every-Product*** method for representing these computations places the result of each product of single-place components on a separate line, being sure to line up the digits in magnitude, just as one does

for addition. The resulting numbers are then added together:

$$\begin{array}{r} 457 \\ \times\ 36 \\ \hline 42 \\ 300 \\ 2400 \\ 210 \\ 1500 \\ \hline 12000 \\ 16452 \end{array}$$

When this method is used in elementary school, it is essential for children to come to understand that, for example, the component product 210 comes from multiplying the 3 in the tens place, hence 30, times the 7 in the ones place. Similarly, the 12000 comes from multiplying the 3 in the tens place, hence 30, times the 4 in the hundreds place, hence 400. Otherwise, teaching this method simply substitutes a less efficient mechanical procedure for a more efficient one.

2D. Division

The interactions between ideas about place value and division illuminate both the process of division and the properties of numbers in base 10 form. In one direction, the idea of division-with-remainder provides a systematic way to find the digits in a number given in base 10 form. In the opposite direction, the effort to find a method for computing the the nonzero single-place components of a quotient in a division problem leads to another interpretation for long division and the uses of base 10 notation for approximation.

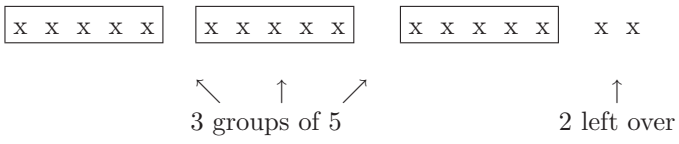
Division-with-Remainder: In its strict sense, division is the inverse operation for multiplication, just as subtraction is the inverse operation for addition. For instance, $6 \div 3 = 2$ *because* $2 \times 3 = 6$ and $72 \div 9 = 8$ *because* $8 \times 9 = 72$. However, division in this strict sense cannot always be carried out within the set of nonnegative integers or even within the set of finite decimal numbers. For example, $3 \div 2$ is not a nonnegative integer (although it can be represented as the finite decimal 1.5) and $5 \div 3$ is not only not a nonnegative integer, but also its decimal representation, 1.666..., is not even finite.

Thus, for the set of nonnegative integers, instead of simply working with division as the inverse of multiplication, we work with *division-with-remainder*, which is a somewhat more general substitute. When division-with-remainder is carried out, instead of yielding a single number as a result, it yields a pair of numbers: a *quotient* and a *remainder*.

For example, when you divide 17 by 5, you get a quotient of 3 and a remainder of 2.

$$\begin{array}{r} 3 \leftarrow \textit{quotient} \\ 5\,\overline{)\,17} \\ \underline{15} \\ 2 \leftarrow \textit{remainder} \end{array}$$

Another way to say this is that 17 equals 3 groups of 5 with 2 left over:[5]



Or,

[5]The result of this division can also be interpreted as 5 groups of 3 with 2 left over, but for simplicity we will stick with the first interpretation.

$$17 = 3 \times 5 \ + \ 2.$$
$$\uparrow \qquad \uparrow$$
3 groups of 5 2 left over

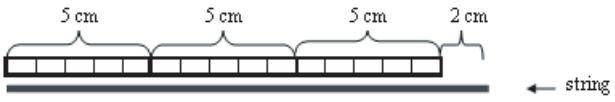
The key consideration here is that the number left over (in this case, 2) should be less than the size of the groups (in this case, 5) because if 5 or more were left over, another group of 5 could be separated off.

One can prove mathematically that this kind of result occurs in general:

> ***Division-with-Remainder Property***: When any nonnegative integer n is divided by any positive integer d, the result is two nonnegative integers: a ***quotient*** q and a ***remainder*** r, where r is smaller than d:
>
> $$n = qd + r \quad \text{and} \quad 0 \leq r < d.$$
>
> Note that either q or r (or both) could equal zero.

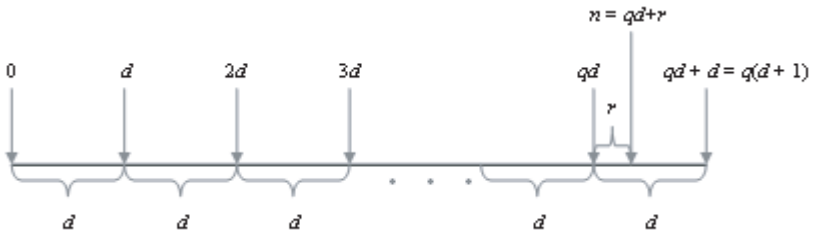
Another way to think of division-with-remainder uses the idea of measurement. For instance, imagine trying to measure a piece of string of length n centimeters using a ruler of length d centimeters and finding that it is possible to fit in q copies of the ruler, but that, when you do this, a piece of string of length r is left over, where r is less than d. Figure 10 shows a picture of that process for $n = 17$ and $d = 5$.

Figure 10: 17 = 5 x 3 + 2



A key property of division-with-remainder is that because $n = qd + r$ and r is nonnegative, n is always between q times d and $(q+1)$ times d. You can visualize this property in general by thinking of a segment of the number line that shows n, some multiples of d, and the remainder r, as shown in Figure 11 Starting at zero and moving q multiples of length d to the right takes you either a little to the left of n (if the remainder r is positive) or directly to n (if the remainder r is zero). But because $r < d$, if you move to the right of qd by one more block of d units you will go past n. So n has to be between qd and $qd+d$, which equals $(q+1)d$.

Figure 11: If $n = qd + r$, then $qd < n < (q + 1)d$



Thus, in general:

> When division of n by d gives a quotient of q and a remainder of r, then
> $$qd \ \leq \ n \ < \ (q+1)d.$$

Division-with-remainder, both theoretically and practically, depends on the order properties of nonnegative integers, which are discussed in section 5. In this section we assume a basic acquaintance with them to give the following formal derivation for the above result: Start with the fact that $n = qd + r$ and the inequality

$$0 \leq r < d.$$

Add qd to all parts of the inequality to obtain

$$qd \leq qd + r < qd + d.$$

Now $qd + r = n$ *(by assumption)* and $qd + d = (q + 1)d$ *(by factoring out d).* So

$$qd \leq n < (q + 1)d.$$

For our example of $n = 17$ and $d = 5$, we have

$$3{\times}5 \leq 3{\times}5 + 2 < 4{\times}5,$$

or, equivalently,

$$3{\times}5 \leq 17 < 4{\times}5.$$

Note that the word quotient is used in two different ways. For instance, we say that the quotient of 11 by 4 is 2 with remainder 3, and we also say that the quotient of 11 by 4 is the number $\frac{11}{4}$, or 2.75. Because using exactly the same term to refer to two different things can cause confusion, when we want to be clear that we are talking about the nonnegative integer q coming from division-with-remainder, we will call it the ***integer quotient*** rather than just the quotient.

Ways to Obtain the Base 10 Form of a Number and Their Interpretations: There are two distinct ways to obtain the base 10 form of a number. The first involves repeated division-with-remainder by 10; the second, though less well known, is important for understanding long division. The second is also fundamental for understanding the approximation properties of base 10 notation (which are discussed in section 5).

***Using Successive Division by 10*:** Observe that the ones digit of a number is the remainder obtained when the number is divided by 10. For example,

$$\text{because} \quad 6,407 = 640 \times 10 + 7, \quad \text{we have that} \quad 6,407 \div 10 \ \text{ is 640 with remainder 7.}$$

Here are the steps showing how the above result was derived:

$6,407$	$=$	$6{\times}1000 + 4{\times}100 + 0{\times}10 + 7$	by converting to expanded form
	$=$	$6{\times}100 \times 10 + 4{\times}10 \times 10 + 0{\times}10 + 7$	because multiplying by 10 adds a zero on the right i.e., by the Law of Exponents: $10^{n-1}{\times}10 = 10^n$
	$=$	$(6{\times}100 + 4{\times}10 + 0){\times}10 + 7$	by the Extended Distributive Rule
	$=$	$640{\times}10 + 7$	by converting from expanded form.

In general:

When a number in base 10 form is divided by 10, the remainder is the rightmost digit of the number and the integer quotient is obtained by taking the number and dropping the rightmost digit.

Thus, if we go back to the example and continue to divide by 10, we obtain the rest of the digits of the number in succession:

$$640 \div 10 \text{ is } 64 \text{ with remainder } 0$$
$$64 \div 10 \text{ is } 6 \text{ with remainder } 4$$
$$6 \div 10 \text{ is } 0 \text{ with remainder } 6.$$

In general:

The digits of a number in base 10 form can be obtained by successive division-with-remainder by 10.

***Using Approximation by a Sequence of Single-Place Numbers*:** A second, seemingly quite different, way to find the composition of a number in base 10 form connects with the idea of approximation and introduces the principle used to perform long division.

Given any number, we can try to approximate it by sums of single-place numbers. Of course there are many ways to do this, but one way is an example of what computer scientists call a ***greedy algorithm*** because at each stage it gets the most it can get at that stage.

Here is the idea of how the greedy algorithm works in this case: Suppose you are given a collection of special numbers, and you want to express an arbitrary number as a sum of the special numbers. For instance, imagine that you want to order a certain number of items from a factory. However, the factory only sells items in full boxes, and it only produces boxes of certain sizes. Thus a procedure is needed to decide which boxes to give you. A greedy algorithm to approximate your order would work as follows: It would first give you the largest possible box: that is, the box containing the largest number of items less than or equal to the number you ordered. Then it would subtract the number of items in that box from the number in your order and give you the largest possible box for the remaining items in your order. It would continue this process as long as possible, each time giving you the largest size box for the remaining unboxed items in your order until either it fills your order completely or there are no available package sizes containing the number that remain in your order or fewer than that number.

A wonderful aspect of the base 10 number system is that if the collection of box sizes is the collection of all the single-place numbers, then the greedy algorithm will always fill your order completely, and it will do so in a way that produces the nonzero single-place components for the number of items in your order. Here is a description of the algorithm:

Approximation Algorithm: Finding the Nonzero Single-place Components of a Number

Step 1: Find the largest single-place number that is less than or equal to the given number.

Step 2: Subtract the number obtained in Step 1 from the given number; the result is the ***error term***.

Step 3: If the error term is greater than 0, go back to Step 1 using the error term in place of the given number.

If the error term is 0, stop: the single-place numbers obtained from the repeated steps of the algorithm are the nonzero single-place components of the original number; their sum is the original number.

As an example, suppose your order consists of 6,407 items. Since

$$6,000 \leq 6,407 < 7,000,$$

the greedy algorithm first gives you a box with $6,000$ items. Then it calculates the number remaining (the first "error term") and finds it to be $6,407 - 6,000 = 407$. So, in the next step, since

$$400 \leq 407 < 500,$$

the algorithm gives you a box with 400 items. Then it calculates the number remaining (the second "error term") and finds it to be $407 - 400 = 7$. So, in the next step, since

$$7 \leq 7 < 8,$$

the algorithm gives you a box with 7 items, which completes your order exactly. Observe that the successive box sizes are precisely the nonzero single-place components of the number of items in your order:

$$6,407 = 7,000 + 400 + 7.$$

The following property of single-place numbers is the reason that the Approximation Algorithm produces the desired result. More precisely, it is the fundamental reason that the sum of single-place numbers produced by the Approximation Algorithm contains at most one term of any given order of magnitude.

> ***Key Property of Single-Place Numbers***: If one starts with any nonzero single-place number and adds the denomination with the same order of magnitude as the number, one obtains the next larger single-place number.

For instance, if one starts with the number 6,000 in the example above and adds the denomination 1,000, one obtains 7,000, which is the next larger number after 6,000 that is a product of a digit times a denomination. Similarly, if one starts with 400 and adds the denomination 100, one obtains 500, which is the next larger single-place number after 400.

In general, when the digit of the single-place number is 1 through 8 and one adds the denomination with the same order of magnitude, one only needs to increase the digit by 1 to obtain the next larger single-place number. (For instance, $60 + 10 = 70$. However, when the digit of the single-place number is 9, the next larger single-place number is obtained by replacing the 9 by a 10. For example, if the single-place number is 9,000, the number 10,000 is the next larger single-place number. Note, however, that the Key Property of Single-Place Numbers holds in this case also because $1,000$ is the denomination with the same order of magnitude as $9,000$ and $9,000 + 1,000 = 10,000$.

We now use the Key Property of Single-Place Numbers to give a formal proof of the fact that the greedy algorithm produces the nonzero single-place components for any number in base 10 form. Let n be any nonnegative integer, and let c be the largest of its single-place components. By the Key Property of Single-Place Numbers, the next larger single-place number is $c + m$, where m is the denomination with the same order of magnitude as c. Since c is chosen to be as large as possible, we know that

$$c \ \leq \ n \ < \ c + m.$$

So, by subtracting c from all three numbers, we have

$$0 \ \leq \ n - c \ < \ m.$$

Now the next stage of the approximation process is to approximate the remainder $n - c$ by its largest single-place component, and since $n - c < m$, the number obtained in this next stage will have a smaller order of magnitude than m. This argument shows that when the algorithm is finally complete, the final sum will be the original number and will have at most one term with any given order of magnitude.

An interesting observation comes out of this way of thinking: If n is a nonnegative integer, c its largest single-place component, and $r = n - c$, then

$$c \ > \ r.$$

Add c to both sides of this inequality to obtain

$$2c \ > \ c + r \ = \ n.$$

Then dividing by 2 gives

$$c > \frac{n}{2}.$$

We may state this result in words:

> For any nonnegative integer n, the largest single-place component of n accounts for more than half of n.

For example, when $n = 199$, then the largest single-place component of n is $c = 100$, half of n is $\frac{n}{2} = \frac{199}{2} = 99.5$, and $100 > 99.5$. In general, however, the largest single-place component of a number is much larger than half the number. For example, the largest single-place component of 985 is 900, which is much larger than half of 985. This topic is discussed in greater detail in section 5.

Approximation and Long Division: In this section we show how the ideas we have been developing underlie the process of long division. The basis for the discussion is that in division-with-remainder of n by d, the integer quotient is the largest nonnegative integer q such that $dq \leq n$. Because the search for an integer satisfying this kind of property formed the basis for each of the steps in the greedy algorithm to find the base 10 expansion of a number, we can adapt the explanation for the success of the greedy algorithm to explain why the long-division process produces the correct answer. In fact, the justification for what we call the Basic Fact of Long Division uses the same Key Property of Single-Place Numbers that guaranteed that the greedy algorithm would produce the nonzero single-place components of a number.

> ***Basic Fact of Long Division***: Suppose that q is the integer quotient of n divided by d. In other words, suppose that $n = qd + r$, where $0 \leq r < d$. Then the largest single-place component of q is the largest single-place number s such that $sd \leq n$.

Before discussing the justification for the Basic Fact of Long Division, we examine how it is used in the long-division process. In general, the problem of long division is to find the base 10 representations for the nonnegative integer quotient q and the remainder r of the division of one nonnegative integer n by a positive integer d. In other words, the problem is to find the base 10 representations for nonnegative integers q and r such that $n = qd+r$ and $0 \leq r < d$. By definition of base 10 notation, this is equivalent to finding (and then adding up) all the single-place components of q for the division of n by d and also to finding the remainder for this division.

Now the Basic Fact of Long Division says that the largest single-place component of q is the largest single-place number s with $sd \leq n$. To find the second largest single-place component of q, which we will call s_1, we subtract sd from n to obtain the difference $n_1 = n - sd$ and repeat the process with n_1 in place of n. Then the Basic Fact of Long Division tells us that s_1 is the largest single-place number with $s_1 d \leq n_1$. Continuing in this way is exactly what one does in long division, until eventually one obtains a difference that is less than the divisor d. This turns out to be the remainder r, and the sum of the single-place components obtained in the individual steps is the integer quotient q. We summarize this procedure in the following algorithm:

Algorithm for the Long Division of n by d

Step 1: Find the largest single-place number s whose product with d is less than or equal to n.

Step 2: Subtract s from n; call the result the ***error term***.

Step 3: If the error term is greater than or equal to d, go back to Step 1 using the error term in place of n.

If the error term is less than d, then it is the remainder of the division of n by d and the sum of the single-place numbers obtained from the repetitions of Step 1 is the integer quotient of the division.

As an example, suppose we want to find the integer quotient and remainder of the division of $n = 763$ by $d = 32$. This is what the computation looks like using one of the standard formats for long division.

$$\begin{array}{r} 23 \leftarrow \textit{quotient} \\ 32\,\overline{)\,763} \\ \underline{64} \\ 123 \\ \underline{96} \\ 27 \leftarrow \textit{remainder} \end{array}$$

The reasoning that explains this process follows the steps of the algorithm:

Step 1: We seek the largest single-place number s whose product with the divisor $d = 32$ is less than or equal to the number $n = 763$. In other words, we want the largest number s with

$$s{\times}32 \;\leq\; 763.$$

We find that $s = 20$ because

$$20{\times}32 \;=\; 640 \;\leq\; 763 \quad \text{whereas} \quad 30{\times}32 \;=\; 960 \;>\; 763.$$

(Thus the largest single-place component of the quotient is 20.)

Step 2: The difference between n and $s{\times}32$ is $763 - 640 = 123$. This is the error term.

Step 3: Since 32 is less than 123, we repeat the process of Step 1 for the number $n_1 = 123$, using the same divisor $d = 32$.

Step 1 (second time): In this step, therefore, we seek the largest single-place number s_1 whose product with 32 is less than or equal to 123. In other words, we want the largest number s_1 with

$$s_1{\times}32 \;\leq\; 123.$$

We find that $s_1 = 3$ because

$$3{\times}32 \;=\; 96 \;\leq\; 123 \quad \text{whereas} \quad 4{\times}32 \;=\; 128 \;>\; 123.$$

(Thus the second largest single-place component of the quotient is 3.)

Step 2 (second time): The difference between n_1 and $s_1{\times}32$ is $123 - 96 = 27$. This is the new error term.

Step 3 (second time): Since $27 < 32$, the error term is less than the divisor 32, and so the process stops. We conclude that the remainder is 27 and the integer quotient is the sum of the single-place numbers obtained in the repetitions of Step 1: $s + s_1 = 20 + 3 = 23$.

Justification for the Basic Fact of Long Division: The justification for the Basic Fact of Long Division parallels the discussion for the approximation interpretation of the base 10 expansion. Suppose s is chosen to be the largest single-place number such that $sd \leq n$ and suppose m is the denomination with the same order of magnitude as s. Then, by the Key Property of Single-Place Numbers, the next larger single-place number after s is $s + m$, and thus

$$n < (s + m)d.$$

In addition, because $qd \leq qd + r$ and $qd + r = n$, we also have that

$$qd \leq n.$$

We may put these two inequalities together to obtain

$$qd \leq n < (s + m)d.$$

Dividing by d yields

$$q < s + m.$$

Hence because $s + m$ is the next larger single-place number after s and because $s + m$ is greater than q, we conclude that s is the largest single-place number less than or equal to q. Thus, by the approximation interpretation for the base 10 expansion, s is the largest single-place component of q. This is exactly the claim of the Basic Fact of Long Division.

Summary: The Basic Fact of Long Division makes long division a practical algorithm although executing it does require fluency with multiplication and subtraction. It also demands some skill in estimation, but what is needed in practice ordinarily comes down to division of two- or three- digit numbers by two-digit numbers. As a multi-step procedure, it can seem somewhat lengthy, but a solid understanding of the underlying principle can give students confidence in the process. Note also that because long division starts out by finding the largest digits of the result, rather than the smallest digits as with the algorithms for the other three operations, one can stop after carrying out the steps to a point where one has an answer that is sufficiently accurate for one's purposes. Section 5 on approximation deepens and extends the ideas in this section.

3. Decimal Fractions

Part of the genius of the place-value system for representing numbers is that the same strategy used to express and compute with nonnegative integers extends with only minor modifications to a much broader class of numbers that involve fractional quantities.

Basic Properties of Fractions: In this section we assume basic familiarity with fractions and their properties. We will use the word ***fraction*** to refer to a symbol $\frac{a}{b}$, where a and b are nonnegative integers and b is not zero, and we call a the ***numerator*** and b the ***denominator*** of $\frac{a}{b}$. The symbol $\frac{a}{b}$ represents a number, namely the result of dividing a by b. We make special note of two properties of fractions.

Rule for Multiplying Fractions: The product of two fractions is the fraction whose numerator is the product of the numerators and whose denominator is the product of the denominators. In symbols: Given fractions $\frac{a}{b}$ and $\frac{c}{d}$,

$$\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}.$$

In particular, if a and b are nonnegative integers and $b \neq 0$, then

$$a \times \frac{1}{b} = \frac{a \times 1}{1 \times b} = \frac{a}{1} \times \frac{1}{b} = \frac{a}{b}.$$

Equivalent Fractions Property: If k is any positive integer, then $\frac{ka}{kb}$ represents the same number as does $\frac{a}{b}$ because $\frac{ka}{kb} = \frac{k}{k} \times \frac{a}{b} = 1 \times \frac{a}{b} = \frac{a}{b}$. Thus,

$$\frac{a}{b} = \frac{ka}{kb}.$$

Definition of Decimal Fraction: A ***decimal fraction*** is any fraction whose numerator is a nonnegative integer and whose denominator is a power of 10. In other words, a decimal fraction is any fraction of the form

$$\frac{a}{10^m} \qquad \text{where } a \text{ and } m \text{ are nonnegative integers.}$$

Recall that $10^0 = 1$. Thus every nonnegative integer can be written as a decimal fraction. For example, $1 = \frac{1}{10^0}$ and $3 = \frac{3}{10^0}$. However, there are many fractions that are not decimal fractions. For instance, $\frac{1}{3}$ cannot be written as a single nonnegative integer divided by a power of 10.[6] It follows that the set of decimal fractions is not closed under division because while both 1 and 3 are decimal fractions, their quotient is not.

Given that not all fractions are decimal fractions and given that not all arithmetic can be performed within the set of decimal fractions, why single them out as special? Two features have led to their widespread use, and the advent of electronic calculators has made these features increasingly important. The first is that the computational algorithms developed for numbers in base 10 form extend with almost no change to computations with decimal fractions. Since arithmetic with general fractions seems to cause trouble for a great many people, the relative simplicity of computing with decimal fractions is very attractive.

The second feature is that any real number can be approximated as closely as one wishes by decimal fractions. This is often expressed by saying that the decimal fractions are *dense* in the set of real numbers. Furthermore, the approximation process is straightforward, in the sense that the same "greedy algorithm," described in section 2D, that works for nonnegative integers applies to general decimal fractions. This combination of familiar, effective computational methods with the ability to approximate means that one can do approximate arithmetic to arbitrary accuracy with arbitrary numbers. So, for practical purposes, working with decimal fractions allows us to compute anything we want. In fact, as will be seen in section 5, the accuracy of decimal approximations increases rapidly with the number of digits used, so that we can usually get enough accuracy without having to do a great many computations.

Multiplication and Addition with Decimal Fractions, I: Given decimal fractions $\frac{a}{10^m}$ and $\frac{b}{10^n}$, we can multiply them by using the Rule for Multiplying Fractions, the Any-Which-Way Rule for multiplication of integers, and the Law of Exponents. The result is as follows:

Formula for Multiplying Decimal Fractions: Given decimal fractions $\frac{a}{10^m}$ and $\frac{b}{10^n}$,

$$\frac{a}{10^m} \times \frac{b}{10^n} = \frac{a \times b}{10^m \times 10^n} = \frac{ab}{10^{m+n}}.$$

To add two decimal fractions $\frac{a}{10^m}$ and $\frac{b}{10^n}$, as for any two fractions, we need to put them over a common denominator. However this is easier for decimal fractions than for general fractions because one of the powers 10^m or 10^n has to divide the other. To be precise, suppose that $m \leq n$. Then by the Law of Exponents, $10^n = 10^m \times 10^{n-m}$. Thus

$$\frac{a}{10^m} = \frac{a \times 10^{n-m}}{10^m \times 10^{n-m}} = \frac{a \times 10^{n-m}}{10^n}.$$

[6]Suppose that $\frac{1}{3} = \frac{a}{10^m}$ for some nonnegative integers a and m. Multiplying both sides by 3×10^m gives $10^m = 3a$, and since 3 is a factor of the righthand side of the equation, it must also be a factor of the lefthand side. But this cannot be the case because 3 is not a factor of 10. So the equation $\frac{1}{3} = \frac{a}{10^m}$ cannot ever be true, no matter what positive integers we might try to substitute for a and m.

The following formula follows immediately:

Formula for Adding Decimal Fractions: Given decimal fractions $\frac{a}{10^m}$ and $\frac{b}{10^n}$ with $m \le n$,

$$\frac{a}{10^m} + \frac{b}{10^n} = \frac{a \times 10^{n-m}}{10^n} + \frac{b}{10^n} = \frac{a \times 10^{n-m} + b}{10^n}.$$

Since we know how to do the integer arithmetic to compute the numerators of the final expressions in the formulas for multiplying and adding decimal fractions, these formulas provide practical and easily applied procedures for doing arithmetic. As an example of how they work, consider multiplying $\frac{638}{10}$ and $\frac{47}{100}$:

$$\frac{638}{10} \times \frac{47}{100} = \frac{638 \times 47}{10 \times 100} = \frac{29,986}{1000}.$$

In addition, because

$$\frac{638}{10} = \frac{638 \times 10}{10 \times 10} = \frac{6380}{100},$$

we have that

$$\frac{638}{10} + \frac{47}{100} = \frac{6380}{100} + \frac{47}{100} = \frac{6380 + 47}{100} = \frac{6427}{100}.$$

Decimal Expansions: As stated above, the formulas for multiplying and adding decimal fractions do not indicate why these fractions have special importance. In fact, we could write similar formulas for fractions having denominators that are not powers of 10. The reason for singling out denominators which are powers of 10 is that we can extend the notion of single-place number to include digits *divided* by powers of 10 as well as digits *multiplied* by powers of 10. With this extension, every decimal fraction can be expressed as a sum of single-place numbers. For example, the numbers used in the preceding calculations can be written as follows:

$$\frac{638}{10} = \frac{600 + 30 + 8}{10} = \frac{600}{10} + \frac{30}{10} + \frac{8}{10} = 60 + 3 + \frac{8}{10}$$

and

$$\frac{47}{100} = \frac{40 + 7}{100} = \frac{40}{100} + \frac{7}{100} = \frac{4}{10} + \frac{7}{100}.$$

This extension of our base 10 notation is further strengthened if we use the standard notation for negative exponents. By definition

$$b^{-n} = \frac{1}{b^n} \qquad \text{for all positive numbers } b \text{ and } n.$$

With this definition the Law of Exponents is valid for all integer exponents:

Law of Exponents: For all integers m, n, and b with $b > 0$,

$$b^m \times b^n = b^{m+n}.$$

By using negative exponents we may rewrite the previous examples as

$$\frac{638}{10} = 6{\times}10^1 + 3{\times}10^0 + 8{\times}10^{-1}$$

and

$$\frac{47}{100} = 4{\times}10^{-1} + 7{\times}10^{-2}.$$

In general, the base 10 system for writing nonnegative integers is based on the denominations

. . .	d_6	d_5	d_4	d_3	d_2	d_1	d_0
. . .	1,000,000	100,000	10,000	1000	100	10	1
. . .	10^6	10^5	10^4	10^3	10^2	10^1	10^0

Extending the system of denominations to include powers with negative exponents gives the following result:

. . .	d_2	d_1	d_0	d_{-1}	d_{-2}	d_{-3}	. . .
. . .	100	10	1	$\frac{1}{10}$	$\frac{1}{100}$	$\frac{1}{1000}$	. . .
. . .	10^2	10^1	10^0	$\frac{1}{10^1}$	$\frac{1}{10^2}$	$\frac{1}{10^3}$	. . .
. . .	10^2	10^1	10^0	10^{-1}	10^{-2}	10^{-3}	. . .

Thus we may extend the definition of ***single-place number*** to be the product of a digit times *any* power of 10 – positive, negative, or zero, and so we now regard all of the following as single-place numbers:

$$4 \times 10^2, \quad 0 \times 10^1, \quad 2 \times 10^0, \quad 5 \times 10^{-1}, \quad 2 \times 10^{-2}, \quad 3 \times 10^{-3}.$$

When n is negative, we may call $a \times 10^n$ a ***fractional single-place number***. The definition of ***order of magnitude*** of a single-place number remains the same: it is the exponent of the power of 10 used to express the number. Thus, for example, $\frac{3}{100}$ has order of magnitude -2 because

$$\frac{2}{100} \;=\; 2 \times \frac{1}{100} \;=\; 2 \times \frac{1}{10^2} \;=\; 2 \times 10^{-2}.$$

More generally, if d is an integer from 1 to 9, then the order of magnitude of $\frac{d}{10^n}$ is $-n$.

Since every nonnegative integer is a sum of single-place components, knowing the corresponding digit for each power of 10 is enough to determine a number's value. This is the information the base 10 notation gives us: it writes the sequence of digits, from right to left, beginning with the digit for $1 = 10^0$. The location of a digit in this sequence tells us the power of 10 by which it should be multiplied.

We now see that decimal fractions have a similar expansion: they can be written as sums whose terms are digits times powers of 10, possibly using negative as well as positive powers. Of course, if we only need to express nonnegative integers in base 10 notation, we do not have to include a decimal point because the power of 10 by which each digit is to be multiplied is clear: the rightmost digit of the number is to be multiplied by 1, the next digit to the left by 10, the one to the left of that by $10^2 = 100$, and so forth.

A given decimal fraction is determined by the digits which multiply the powers of 10, noting that these digits are only nonzero for finitely many powers. The digits naturally make a sequence, with the position of a digit in the sequence determined by which power of 10 it multiplies. When we were dealing with integers, this sequence had a natural start, namely the ones place (order of magnitude zero), and progressed indefinitely far to the left. With fractions, however, the sequence can extend arbitrarily far in both directions. So, to know which digits go with which powers of 10, we have to show explicitly where to find the location of the digit with magnitude zero (the ones place).

The standard way to do this is to place a mark, the decimal point, between the digit for $10^0 = 1$ and the digit for $10^{-1} = \frac{1}{10}$. The digits to the left of the decimal point are multiplied by nonnegative powers of 10, and the digits to its right are multiplied by successive negative powers of 10. For the numbers in the previous example, we have

$$\frac{638}{10} = 60 + 3 + \frac{8}{10} = 63.8, \qquad \text{and} \qquad \frac{47}{100} = \frac{4}{10} + \frac{7}{100} = .47.$$

This representation allows us to write the number compactly in ***(extended) base 10 notation*** as

$$d_n d_{n-1} \cdots d_2 d_1 d_0 \, . \, d_{-1} d_{-2} \cdots d_{-m}.$$

With this notation we can easily see why the usual procedure for multiplying numbers written in extended base 10 notation works as well as it does. The crucial fact is that the number of places to the right of the decimal point equals the power of 10 in the denominator of the fraction. Thus to multiply two numbers, one can first ignore the decimal points and multiply the numbers as if they were integers. Then to place the decimal point in the product, one can use the following rule: If the decimal point is m places to the the left of the rightmost digit in the first number and n places to the left of the rightmost digit in the second number, then the decimal point in the product is $m+n$ points to the left of the rightmost digit in the product. This procedure follows directly from the fact that in the formula for multiplying decimal numbers the power of 10 in the denominator is the sum of the powers of 10 in the two factors. For example:

$$63.8 \times .47 \;=\; \frac{638}{10^1} \times \frac{47}{10^2} \;=\; \frac{638 \times 47}{10^{1+2}} \;=\; \frac{29,986}{10^3} \;=\; 29.986$$

In addition, note that multiplying by 10 is extremely easy to do. It simply turns each denomination into the next larger one, and so each digit is multiplied by the next larger denomination and, therefore, gets shifted one place to the left relative to the decimal point. The result is that in the number as a whole, the decimal point is shifted one place to the right. Similarly, division by 10 simply shifts the decimal point one place to the left. For example,

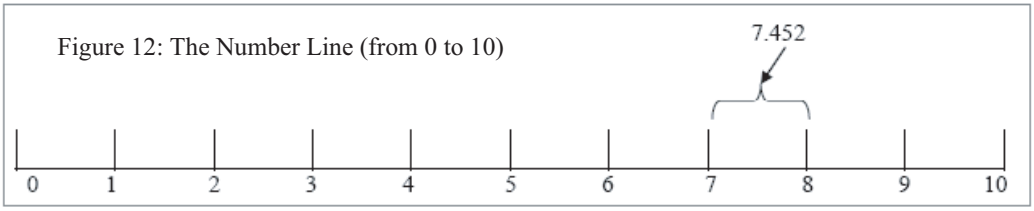
$$74.52 \times 10 \;=\; 745.2 \qquad \text{and} \qquad 74.52 \div 10 \;=\; 7.452.$$

Decimal Expansions as an Address System on the Number Line A good way to think about the decimal expansion is that it gives an address system for locating numbers on the number line. This gives a geometric way to picture how a number is approximated by the sum of its leading decimal components. Consider a number like

$$7.452 \;=\; \frac{7452}{1000} \;=\; 7 + \frac{4}{10} + \frac{5}{100} + \frac{2}{1000} \;=\; 7 + .4 + .05 + .002.$$

Let's locate it on the number line. Here is a piece of the number line that goes from 0 to 10:

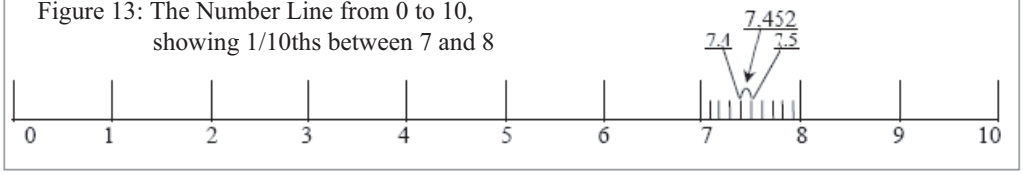
Figure 12: The Number Line (from 0 to 10)



The 7 in 7.452 tells us that that the number is is between 7 and 8 on the number line.

To locate 7.452 more precisely, we look at the next place in its decimal expansion. We divide the interval between 7 and 8 into ten equal subintervals.

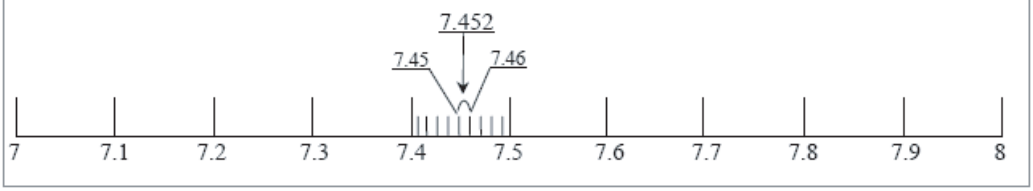
Figure 13: The Number Line from 0 to 10, showing 1/10ths between 7 and 8



The 4 in 7.452, which represents $\frac{4}{10}$, or $4 \times \frac{1}{10}$, tells us that the number is located between 7.4 and 7.5 on the number line.

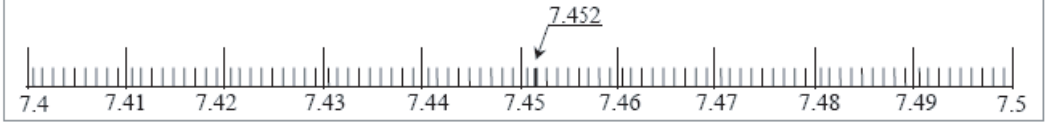
To locate 7.452 even more accurately, we divide the interval between 7.4 and 7.5 into 10 equal subintervals, each of which is one-hundredth of a unit wide. In order to show this, we magnify the interval between 7 and 8.

Figure 14: The Number Line from 7 to 8, showing 1/10ths between 7 and 8, and 1/100ths between 7.4 and 7.5



As we proceed, the divisions become finer and finer, in fact they become almost too fine for us to see the subintervals! This is graphic proof of the rapidity with which the decimal expansion approximates a number. If, however, we magnify the interval between 7.4 and 7.5, then we can see what happens at the next step.

Figure 15: The Number Line from 7.4 to 7.5



Notice that when we expand the interval from 7.4 to 7.5 by a factor of 100, it looks just like the interval from 0 to 10, except that the labels are changed. The division points 7.40, 7.41, 7.42, etc., are seen to divide the interval into 10 equal subintervals. The next single-place component, namely the 5 in 7.45, tells us that 7.452 is in the subinterval between 7.45 and 7.46. Finally, the last decimal component, the 2, locates 7.452 in one of the 10 equal subintervals between 7.45 and 7.46. We are now working at distances which we cannot distinguish visually. We would need a microscope to really see what is going on.

For practical purposes, knowing four terms of the decimal expansion of a number locates it to a high degree of precision. However, in principle, decimal fractions can be located to any number of places. In other words, because of the arbitrary fineness of the divisions produced by the decimal fractions, any decimal fraction can be located with arbitrary accuracy on the number line, by a decimal address, using the procedure described above. More generally, given any real number, whether a decimal fraction or not, we can approximate the number by a decimal fraction. For example the number 7.452 serves as an approximate address for any real number in the interval between 7.452 and 7.453. This is essentially the idea behind using significant digits to record "real life" numbers.

Multiplication and Addition with Decimal Fractions, II: The previous examples of adding and multiplying $\frac{638}{10} = 63.8$ and $\frac{47}{100} = .47$ made use of the formulas for adding and multiplying decimal fractions, but they did not include the steps required to multiply and add the integer numerators that occurred in the calculations. The fact that decimal fractions can be written as sums of expanded single-place numbers, along with the facts that the Rules of Arithmetic apply to decimal fractions as well as to integers, allow us to perform computations similar to the ones we use for integers directly with the decimal fractions. For

example:

$$
\begin{aligned}
63.8 + .47 &= \frac{638}{10} + \frac{47}{100} && \text{by converting to fractional notation} \\
&= (60 + 3 + \frac{8}{10}) + (\frac{4}{10} + \frac{7}{100}) && \text{by converting to expanded form} \\
&= 60 + 3 + (\frac{8}{10} + \frac{4}{10}) + \frac{7}{100} && \text{by the Any-Which-Way Rule} \\
&= 60 + 3 + \frac{12}{10} + \frac{7}{100} && \text{by adding components of the same magnitude} \\
&= 60 + 3 + (1 + \frac{2}{10}) + \frac{7}{100} && \text{by converting to expanded form} \\
&= 60 + (3 + 1) + \frac{2}{10} + \frac{7}{100} && \text{by the Any-Which-Way Rule} \\
&= 60 + 4 + \frac{2}{10} + \frac{7}{100} && \text{by adding components of the same magnitude} \\
&= \frac{6427}{100} && \text{by converting from expanded form} \\
&= 64.27 && \text{by converting to base 10 notation.}
\end{aligned}
$$

Observe that at each stage of the computation only components of the same magnitude are added. This corresponds exactly to the usual procedure for adding two decimal numbers: Line up the decimal points, place the decimal point for the answer directly underneath them, and add the two numbers using exactly the same technique as one would if they were nonnegative integers:

$$
\begin{array}{r}
63.8 \\
+\ \underline{.47} \\
64.27 \\
\uparrow \\
\text{keep the decimal points lined up}
\end{array}
$$

The factor of 10^{n-m} in the formula for adding decimal fractions corresponds to the lining up of decimal points in this computation. Note that for mental mathematics, if one would be satisfied with an approximate answer, one could start at the lefthand side, adding the digits of greatest magnitude, and stop after a few places.

In the example of multiplying $\frac{638}{10}$ and $\frac{47}{100}$, we also omitted the computations for multiplying the numerators of the fractions and simply wrote that $638 \times 47 = 29,986$. The computations used to multiply $\frac{638}{10}$ and $\frac{47}{100}$ are very similar to those used to multiply 638 and 47:

$$
\begin{aligned}
63.8 \times .47 &= \frac{638}{10} \times \frac{47}{100} && \text{by converting to base 10 form} \\
&= (60 + 3 + \frac{8}{10}\ ) \times (\frac{4}{10} + \frac{7}{100}) && \text{by converting to expanded form} \\
&= (60\times\frac{4}{10}) + (60 \times \frac{7}{100}) + (3\times\frac{4}{10}) + (3\times\frac{7}{100}) + (\frac{8}{10}\times\frac{4}{10}) + (\frac{8}{10} \times\frac{7}{100}) \\
& && \text{by the Extended Distributive Rule} \\
&= \frac{240}{10} + \frac{420}{100} + \frac{12}{10} + \frac{21}{100} + \frac{32}{100} + \frac{56}{1000} \\
& && \text{by the Any-Which-Way Rule} \\
&= (20 + 4) + (4 + \frac{2}{10}) + (1 + \frac{2}{10}) + (\frac{2}{10} + \frac{1}{100}) + (\frac{3}{10} + \frac{2}{100}) + (\frac{5}{100} + \frac{6}{1000}) \\
& && \text{by converting to expanded form} \\
&= 20 + (4 + 4 + 1) + (\frac{2}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10}) + (\frac{1}{100} + \frac{2}{100} + \frac{5}{100}) + \frac{6}{1000} \\
& && \text{by the Any-Which-Way Rule} \\
&= 20 + 9 + \frac{9}{10} + \frac{8}{100} + \frac{6}{1000} && \text{by adding terms of the same magnitude} \\
&= 29.986 && \text{by converting to base 10 form.}
\end{aligned}
$$

4. Base 10 Components and Algebra

The Role of the Rules of Arithmetic in Algebra: There are several ways in which having students make explicit use of single-place numbers in arithmetic can help prepare them for algebra. One is that, as they develop facility in applying the Rules of Arithmetic to solve problems with them, they are, in effect, practicing the same computational techniques they will use in algebra. The reason is that the Rules of Arithmetic are also the rules for manipulating algebraic expressions. The Rules provide a remarkably compact summary for much of what constitutes legitimate algebraic manipulation. Moreover, they themselves are most succinctly expressed as algebraic identities. So, for students who have developed an increased awareness of them, writing them symbolically can become part of a natural transition to the use of algebraic notation for expressing general properties.

Base 10 Form and Polynomials in 10: Working with base 10 components can also prepare students for algebra through the close relationship between decimal arithmetic and the algebra of polynomials. In a certain sense base 10 notation exploits algebra in the service of arithmetic because decimal numbers can be usefully thought of as "polynomials in 10." Emphasizing this relationship can both shed light on arithmetic and make algebra more familiar and learnable.

For example, consider the numbers 21 and 13 and the expressions $2x + 1$ and $x + 3$, and observe the similarity between the computations for $21 + 13$ and $(2x + 1) + (x + 3)$, with the symbol x playing the role of the number 10 and using juxtaposition of symbols to represent multiplication in algebraic expressions.

Line 1	$21 + 13$	
Line 2	$= (20 + 1) + (10 + 3)$	$(2x + 1) + (x + 3)$
Line 3	$= (2\times10 + 1) + (10 + 3)$	$= (2x + 1) + (x + 3)$
Line 4	$= (2\times10 + 10) + (1 + 3)$	$= (2x + x) + (1 + 3)$
Line 5	$= (2\times10 + 1\times10) + 4$	$= (2x + 1x) + 4$
Line 6	$= (2 + 1)\times10 + 4$	$= (2 + 1)x + 4$
Line 7	$= 3\times10 + 4$	$= 3x + 4$
Line 8	$= 30 + 4$	$= 3x + 4$
Line 9	$= 34$	

Note that in both cases, lines 1–3 translate the expression into a less compact form, making the implicit multiplications and additions explicit. There is one more step in the case of $21 + 13$ because the addition of terms is already explicit in the polynomial, but for 21 and 13 they are made explicit through use of the definition of base 10 notation. The Any-Which-Way Rule for Addition justifies the step from line 3 to line 4, the fact that 1 times any number equals the number justifies the step from line 4 to line 5, and the Distributive Rule justifies the step from line 5 to line 6. The last lines recompactify the expressions, using the definition of base 10 notation in the case of the numerical problem.

The computations for 21×13 and $(2x + 1)(x + 3)$ also parallel each other, again with x playing the role of 10.

Line 1	21×13	
Line 2	$= (20 + 1) \times (10 + 3)$	$(2x + 1)(x + 3)$
Line 3	$= (2 \times 10 + 1) \times (10 + 3)$	$= (2x + 1)(x + 3)$
Line 4	$= (2 \times 10) \times 10 + (2 \times 10) \times 3 + 1 \times 10 + 1 \times 3$	$= (2x)x + (2x)3 + 1x + (1)(3)$
Line 5	$= 2 \times (10 \times 10) + (2 \times 3) \times 10 + 10 + 3$	$= 2(xx) + (2)(3)x + x + 3$
Line 6	$= 2 \times 100 + 6 \times 10 + 1 \times 10 + 1$	$= 2x^2 + 6x + 1x + 3$
Line 7	$= 200 + (6 + 1) \times 10 + 3$	$= 2x^2 + (6 + 1)x + 3$
Line 8	$= 2 \times 10^2 + 7 \times 10 + 3$	$= 2x^2 + 7x + 3$
Line 9	$= 273$	

As in the previous example, lines 1–3 translate the expression into a less compact form. The Extended Distributive Rule justifies the step from line 3 to line 4 and the Any-Which-Way Rule for Multiplication justifies the step from line 4 to line 5, the fact that 1 times any number equals the number justifies the step from line 5 to line 6, and the Distributive Rule justifies the step from line 6 to line 7. Finally, as in the previous example, the last lines recompactify the expressions.

Further investigation reveals that the analogy between the two types of computations is not absolutely perfect although it remains useful. For example $(2x+4)+(3x+7) = 5x+11$, whereas $24+37 = 61$. However, if we plug $x = 10$ into $5x + 11$, we get $50 + 11$ and $50 + 11$ is the intermediate result obtained in adding $24 + 37$, before ten of the 1s are combined into a 10 to get the standard decimal form of $50 + 11 = 61$. In other words, polynomial arithmetic is like decimal arithmetic without the regrouping process. It follows that the coefficients of the terms in a polynomial may become arbitrarily large.

Another analogy between polynomial arithmetic and decimal arithmetic is in the relation between the degree of a polynomial and the order of magnitude of a decimal number. Recall that the degree of a polynomial is the highest power of x that appears with a non-zero coefficient in a term of the polynomial. Although this analogy is not perfect, it is very helpful. For example, just as order of magnitude gets us started when we divide numbers in base 10 form, degree is what we begin with to find the quotient of two polynomials. In addition, the recursive procedure for finding the quotient of two polynomials is very similar to the one previously described above for long division of numbers in base 10 form. Thus, preparing students for polynomial division in algebra is one reason to encourage them to become comfortable with long division of numbers.

Because of the absence of carrying in polynomial arithmetic, degree is actually easier to work with than order of magnitude in a number of ways. For instance, the degree of the sum of two polynomials is always less than or equal to the larger of their degrees, whereas the order of magnitude of a sum of numbers can be larger than either order of magnitude of the addends. (As an example, both 882 and 755 have order of magnitude 2, but their sum is 1,637, which has order of magnitude 3.) Moreover, the degree of a product of two non-zero polynomials is exactly the sum of the degrees of the two factors, whereas the order of magnitude of a product of two numbers can be larger than the sum of the orders of magnitude of the factors. (As an example, both 69 and 57 have order of magnitude 1, and so the sum of their orders of magnitude is 2. But their product is 3,933, which has order of magnitude 3.)

Identities in Arithmetic: Besides the identities that describe the Rules for Arithmetic, there are other important identities that can be used to illustrate connections between arithmetic and algebra. For example, the identity for the difference between the squares of two numbers

$$x^2 - y^2 = (x + y)(x - y) \qquad \text{for all real numbers } x \text{ and } y$$

can be used to do mental arithmetic, such as computing the product of 38 and 42.

$$\begin{aligned} 38 \times 42 &= (40 - 2) \times (40 + 2) \\ &= 40^2 - 2^2 \\ &= 1600 - 4 = 1596. \end{aligned}$$

If the identity is rewritten by adding y^2 to both sides, it becomes

$$x^2 = (x + y)(x - y) + y^2$$

and can be used to compute a quantity like 45^2:

$$\begin{aligned} 45^2 &= (45 - 5) \times (45 + 5) + 5^2 \\ &= 40 \times 50 + 5^2 \\ &= 2000 + 25 = 2025. \end{aligned}$$

Other versions of the identity can be used to develop additional mental arithmetic skills and to find factorizations of numbers whose prime factors are greater than 20. The factoring of large numbers is a computation of great importance in connection with contemporary computer security algorithms.

On Beyond the Basics: Regarding decimal numbers as "polynomials in 10" amounts to giving the variable x the value 10. This is called ***specialization***. The specialization that turns polynomial arithmetic into decimal arithmetic requires, in effect, that x satisfy the equation $x = 10$, or, equivalently, $x - 10 = 0$. Specialization using polynomials can also be used to interpret number systems with bases other than 10 by requiring x to satisfy other equations. For example, requiring x to satisfy the equation $x^2 - 2 = 0$ means that x acts like the irrational number $\sqrt{2}$ and makes polynomials act like numbers of the form $a + b\sqrt{2}$. Requiring x to satisfy $x^2 + 1 = 0$ means that x acts like the imaginary number i and polynomials act like numbers of the form $a + bi$. Thus, in this case, the polynomials act like complex numbers. These facts show how ramifications of the idea of place value extend to the very end of the K-12 mathematics curriculum. In fact, they extend even beyond because the idea of specialization and its consequences form an important theme in abstract algebra.

5. Single-place Numbers, Estimation, and Error

Order and Magnitude: Besides the arithmetic operations, the other main structure on the whole numbers is order:

We define one number to be larger than another number to mean that the first equals the second plus a positive number. In symbols:

$$a > b \text{ means that } a = b + c, \text{ where } c \text{ is a positive real number.}$$

The decimal representation of numbers makes comparison of sizes among numbers quite simple. A basic fact is that order of magnitude sorts positive numbers according to size: *any* positive number of a given order of magnitude is larger than *any* other of a smaller order of magnitude. For instance, 1000, the smallest positive number of magnitude 3, is greater than every number of magnitude 2, even those very close to 1000, such as 999.99999. This simple fact has several important consequences. The first is a recipe for comparing any two numbers in base 10 form.

Ordering Algorithm: If a and b are two numbers in base 10 form, to decide which is larger compare their single-place components. Find the largest order of magnitude for which the single-place components of a and b are different. Then the number with the larger component of that magnitude is larger.

In particular, if the order of magnitude of a is larger than the order of magnitude of b, then a is larger than b. If they have the same order of magnitude but the largest single-place component of a is larger than the largest single-place component of b, then a is larger. If the largest single-place components agree, proceed downward in orders of magnitude, until you first find a magnitude at which the components of a and b differ. Then the larger number is the one with the larger component of that magnitude.

Expressed in terms of base 10 notation, this means that if a and b have the same order of magnitude but the leading digit of a is larger than the leading digit of b, then a is larger than b. If both a and b have several identical leftmost digits and the first non-identical digit of a, reading from left to right, is larger than the corresponding digit of b, then a is larger than b.

Relative Place Value: Often, we want to do more than say whether one number is larger than another, we want to say how much larger it is, or that it is very much larger. We also want to say when two numbers are close to one another. Paying attention to single-place numbers also makes this easy to do. The main point is that multiplying by 10 just increases the order of magnitude by 1. This is true no matter what decimal place we are talking about. Thus, given any decimal place, the place just to the left represents numbers 10 times larger than those in the given place, and the place just to the right represents numbers only $\frac{1}{10}$ the size of the given place. The decimal place that is two places to the left of a starting place represents numbers that are 100 times larger than those in the starting place, and the place that is two places to the right represents those only $\frac{1}{100}$ of the size of those in the starting place. For facility with estimation, it is important that students understand not only the values of each place, but also the *relative* values of the places.

Approximation and Error: Relative place value ideas help shed light on the fact that, for numbers of all magnitudes, it is the largest few single-place components (the leftmost few decimal places) that account for most of the number's size. This idea is particularly important in dealing with error, which is unavoidable whenever we must deal with "real-life" numbers. The reason is that real-life numbers come from measurement and are, therefore, known only approximately. The question of how accurately they are known is fundamental. Here the key idea is *relative error*: the size of the error involved in approximating a given number in comparison to the number itself. For instance, if you only have a few dollars, you care quite a bit whether it is \$4 or \$6. However, if you have around a thousand dollars, you probably don't care too much if it is \$1004 or \$1006. And if you have a million dollars, even a thousand dollars more or less, let alone \$2, will probably not keep you up at night.

There are many situations in which we would like to work with a certain number, but because we do not know its precise value or simply because it is more convenient, we use a somewhat different number instead. For example, because π is irrational (and is therefore not equal to a fraction or to a terminating decimal number), we often use the fraction $\frac{22}{7}$ or the decimal number 3.14 in place of π. We could also use more accurate approximations such as $\frac{355}{113}$ or 3.14159. The concept of relative error helps us evaluate the usefulness of an approximation.

Definition: Suppose v is an approximation to V. The ***absolute error*** of approximating V by v is the absolute value of the difference, $|V - v|$. The ***relative error*** of approximating V by v is the ratio $\frac{|V-v|}{V}$, which is the ratio of the absolute error to the correct amount.

This definition means that we can think of relative error as measuring error in the most relevant units, namely units of the correct amount.

Both absolute error and relative error are controlled easily in terms of single-place numbers. The main observation is that, the more alike the decimal expressions (read from the left) of two numbers are, the smaller the bound on their relative difference. To be precise about this, we will suppose that we have two numbers V and v, of the same order of magnitude, and that either their largest single-place components

are equal, or their two largest single-place components, or perhaps even more of them. There are two ways of formulating this: we can specify the number of single-place components that agree, or we can specify the order of magnitude of the largest single-place components that differ. It is easy to convert one type of information into the other, but we distinguish them because they reflect the two points of view - relative versus absolute error.

We can use the decompositions afforded by division-with-remainder by some power of 10, as described in section 2D on division, to describe the relationship between V and v. We suppose that their common order of magnitude is m, and that their single-place components of magnitude greater than or equal to ℓ agree. This means that, if we use division-with-remainder by 10^ℓ on both of them, the integer quotients will be the same. In other words, we can write

$$V = q10^\ell + R \qquad \text{and} \qquad v = q10^\ell + r.$$

The common integer quotient q will have order of magnitude $m - \ell$, which means that it has $m - \ell + 1$ decimal places. The remainders R and r have order of magnitude less than ℓ. In other words, they are less than 10^ℓ. As an example, let $V = 7452$ and $v = 7420$. Then $m = 3$ and $\ell = 2$. Now $10^\ell = 10^2 = 100$, and $V = 74 \times 100 + 52$ and $v = 74 \times 100 + 20$. So the integer quotient of V by 10^ℓ is 74 with a remainder $R = 52$, the integer quotient of v by 10^ℓ is 74 with a remainder $r = 20$, and both R and r have magnitude $m - \ell = 3 - 2 = 1$. In other words, the remainders are both less than $10^\ell = 10^2 = 100$, and, therefore, $|V - v| = |R - r| = |7452 - 7420| = 32$ is less than 100.

In this situation, the following statement holds:

Decimal Estimation Theorem: Suppose that V and v are two numbers of magnitude m, and suppose their single-place components of magnitude ℓ or larger are equal. (That is, the largest $m - \ell + 1$ single-place components of V are all equal to the corresponding components of v.) Let q be their common integer quotient when division-with-remainder by 10^ℓ is used on them. Then

(i) The absolute error $|V - v|$ has order of magnitude at most $\ell - 1$, and hence is less than 10^ℓ.

(ii) The relative error $\dfrac{|V - v|}{V} < \dfrac{1}{q}$. In particular, regardless of q, it is always less than $\dfrac{1}{10^{(m-\ell)}}$.

An important case of the Decimal Estimation Theorem is *rounding down*, which is sometimes called *front-end estimation*. This involves ignoring, or more precisely, replacing by zero, the smaller single-place components of a number. For example, we might approximate 7452 by 7450, or by 7400, or by 7000, depending on the relative importance we place on accuracy versus simplicity.

The relative error of approximating 7452 by 7450 is

$$\frac{|7452 - 7450|}{7452} = \frac{2}{7452} < \frac{10}{7452} < \frac{1}{745} \cong 0.13\%.$$

The last amount is the estimate given by the Decimal Estimation Theorem, with $m = 3$ and $\ell = 1$. The reason the actual relative error is so much smaller than the estimate provided by Decimal Estimation Theorem is because the actual ones digit is 2, but it potentially could be as large as 9, and the Decimal Estimation Theorem must work for all possible ones digits.

The relative error of approximating 7452 by 7400 is

$$\frac{|7452 - 7400|}{7452} = \frac{52}{7452} < \frac{100}{7452} < \frac{1}{74} \cong 1.4\%.$$

The last amount is the estimate provided by the Decimal Estimation Theorem when $m = 3$ and $\ell = 2$. Finally, if we replace 7452 by 7000, the relative error is

$$\frac{|7452 - 7000|}{7452} = \frac{452}{7452} < \frac{1000}{7452} < \frac{1}{7} \cong 14\%,$$

where the final amount is again provided by Decimal Estimation Theorem, with $m = 3$ and $\ell = 3$.

Limits of Accuracy in Practice: The Decimal Estimation Theorem expresses in a concise way the sense in which knowing the beginning (the largest several single-place components) of the decimal expansion of a number tells us most of what we want to know about the number as a magnitude. In certain branches of mathematics, it is important to know all the single-place components of a number, in order to tell, for example, if it is a perfect square, or if it is divisible by 7, or if it is prime. But if it is a number coming from a measurement, then all we need to know, or can expect to know, is an approximate value, and the largest single-place components supply this with great efficiency.

Significant Figures: The term significant figures is used in connection with numbers that result from measurements. To say that a number is known to ***k significant figures*** means that we we have exact values for the largest k single-place components of the number. Part (ii) of the Decimal Estimation Theorem tells us that if we know a number to k significant figures, then the relative error between the measured values and the "true" value is at most $\frac{1}{10^{k-1}}$. So for example, if we know a number to 4 significant figures, we know it with a relative error of less than $\frac{1}{1000}$.

In fact, it is rather rare to know a measured number with such accuracy. Take the example of the "radius" of Earth. It is approximately 4,000 miles, but sometimes it is given as 3,928 miles. However, the last digit does not have a clear meaning. In speaking of the "radius" of Earth, we are pretending that Earth is a perfect sphere. But it is not. It deviates from being a perfect sphere in three ways. First, because of its rotation, it is slightly oblate – flattened at the poles, and thickened around the equator. Second, it has bumps and dimples like Mount Everest and the Challenger Deep. Third, because of the motion of its liquid interior, it is slightly deformed, with a bulge in the north Pacific. These imperfections mean that it does not make sense to speak of a "radius" of Earth more accurately than to about 10 miles. Thus, the radius of the Earth is a number defined only to 3 significant figures. Most other "real-life" numbers have similar limitations. The results of polls are typically accurate only to about $\pm 3\%$, which is slightly better than one significant figure. Although the U.S. Census reports state populations as exact numbers of people, in the millions (6 significant figures) or the tens of millions (7 significant figures), it is lucky if these numbers are accurate to 3 significant figures.

In summary, the notion of significant figures successfully captures many of the key notions of error and approximation for decimal numbers. It especially provides an easy way to deal with relative error. For many purposes, it suffices to know a number to one significant figure. For most purposes, two significant figures are enough, and it is rare to know a "real-life" number to more than 3 significant figures. Numbers used in finance can be exceptions to these rules, but even these are sometimes known only approximately.

Scientific notation: Scientists and others who use numbers resulting from measurements are mainly interested, first, in their size, and second, in the accuracy to which they are known. The way of writing numbers known as scientific notation is designed to exhibit these two aspects very clearly.

Scientific notation is actually an alternative way to write decimal fractions. Instead of writing a decimal fraction d in the conventional way, with a certain number of places to the left of the decimal point, and a certain number to the right, ***scientific notation*** rewrites it in the form

$$d = (\frac{d}{10^m}) \times 10^m,$$

where m is the order of magnitude of d. The result is that $\frac{d}{10^m}$ is a number between 1 and 10 and the magnitude of d is the exponent of 10. Thus, scientific notation separates out a key feature of a number, namely its magnitude, and displays it prominently.

A further convention frequently used by scientists is to report only significant figures when recording a number in scientific notation, so as not to be misleading about the accuracy to which the number is known. Scientists also ordinarily present measurements rounded to the nearest significant digit, rather than rounded down, as assumed in the Decimal Estimation Theorem. By this convention, 2.3×10^4 means a number known to be between 22,500 and 23,500, while 2.30×10^4 means a number known to be between 22,950 and 23,050. With this understanding, scientific notation uses exponents and the idea of significant digits

to express in a very direct way the size and accuracy of a "real-world" number). The Decimal Estimation Theorem, modified slightly to account for the rounding convention usually used in practice, tells us the maximum possible relative error in a number with given significant figures.

Single-place Numbers and Estimation in Addition, Subtraction, and Multiplication: Thinking in terms of relative place value also provides efficient ways of estimating sums and products. To keep the discussion short we will avoid exact details here, but the principle is clear: when estimating, focus on the largest single-place components. Based on an understanding of how single-place numbers enter into arithmetic, we can describe what to do to get an estimate with desired accuracy for a sum or product.

Estimating sums To estimate a sum, simply compute the sums of the largest single-place components. The three largest components of the larger summand, and the components of the same magnitudes for the smaller summand, will give an approximation to the sum with relative error small enough for most purposes. For example, suppose we approximate $83,244$ by $83,200$ and $5,293$ by $5,200$. Then

$$83,244 + 5,293 \cong 83,200 + 5,200 \cong 88,400.$$

In this case the exact sum is $88,537$, and so the relative error is

$$\frac{|88,537 - 88,400|}{88,537} = \frac{137}{88,537} < 0.2\%$$

At its worst, absolute error adds under addition. That is, if u is an approximation to U, with absolute error $|U-u|$, and v is an approximation to V, with absolute error $|V-v|$, then the absolute error of $u+v$ as an approximation to $U+V$ is at most the sum of the absolute errors. This follows from the Triangle Inequality for absolute value:

$$|(U+V)-(u+v)| \;=\; |(U-u)+(V-v)| \;\leq\; |U-u| + |V-v|\ .$$

On the other hand, if all the addends are positive, then it can be shown that the relative error does not increase at all under addition. In other words, if U and V are positive, and u approximates U with a certain relative error, and v approximates V with (at most) the same relative error, then $u+v$ approximates $U+V$ with at most that same relative error.

Estimating differences: When approximate values are used in a subtraction problem, the relative error is quite different from the relative error for addition. In fact, in subtracting one positive number from another, one may completely lose control over relative error: it can become arbitrarily large. This is because the difference of two large numbers may be quite small, while the difference between the approximations for the numbers may be relatively large. The consequence is that the relative error can be extremely large because it is obtained by dividing the absolute error of the difference by the actual value of the difference. For example, if $U = 100.1$ is approximated by $u = 100$ and if $V = 99.9$ is approximated by $v = 99$, then the difference between the true values is $|U-V| = 0.2$ and the difference between the approximate values is $|u-v| = 1$. Thus, the

$$\begin{aligned}\text{relative error} \quad &= \quad \left|\frac{\text{difference between the true values} - \text{difference between the approximate values}}{\text{difference between the true values}}\right| \\ &= \quad \left|\frac{0.2-1}{0.2}\right| \;=\; \frac{0.8}{0.2} \;=\; 4 \;=\; 400\%.\end{aligned}$$

Even more dramatically, if we take two different approximations u_1 and u_2 to the same number U and use the difference $u_1 - u_2$ as an approximation to $U - U = 0$, then as long as u_1 is not equal to u_2, the relative error of the approximation will be infinitely large.

More generally, if u and v are approximations of U and V, each with absolute error e or less, and if the difference between U and V is also less than e, then the number $u - v$ will have absolute value not more

than $2e$. Thus, if we approximate the difference $U - V$ by $u - v$, we have an upper bound on the size of the absolute error but essentially no control over the size of the relative error. For this reason, it is often unwise to use $u - v$ in a further calculation involving $U - V$, particularly one that involves dividing by $U - V$. In less extreme situations, one may retain some information about $U - V$, but with a substantial loss of significant digits.

Estimating products: Recall that a product of decimal numbers is the sum of all the products of the single-place components of one factor times the single-place components of the other factor. The largest of these addends is the product of the largest single-place components of each factor. The next larger are the two terms obtained by taking the product of the second largest component of one factor with the largest component of the other. To illustrate this, look again at the example from section 2C, the product 36×457. The largest product of single-place components is $12,000 = 30 \times 400$. The next two larger products of single-place components are $30 \times 50 = 1500$ and $6 \times 400 = 2400$. The sum of these three terms is $15,900$. If we use this number as an estimate for the product, which is $16,452$, the relative error is

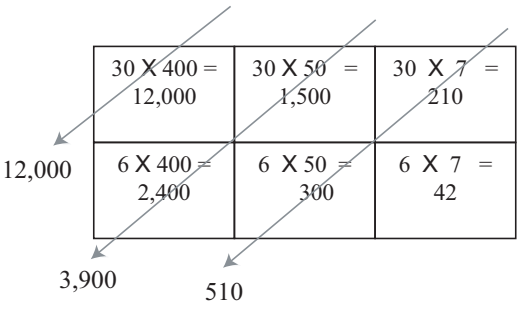
$$\frac{16,452 - 15,900}{16,452} < 4\%.$$

For even more accuracy, we could compute the products of the next smaller single-place components, namely $30 \times 7 = 210$ and $6 \times 50 = 300$. Adding these to the $15,900$ gives $16,410$ and reduces the relative error to

$$\frac{16,452 - 16,410}{16,452} < 0.3\%.$$

Figure 16 shows an attractive way to visualize this situation, which uses the lattice model for multiplication. Imagine the products of the single-place components of the two numbers arranged in an array of boxes, as was done in section 2C. The largest product is in the upper left-hand box. The next larger products are on the diagonal adjacent to this box. The products next in size are located on the next diagonal down to the right, and so forth. Although a full discussion of relative error for multiplication is too technically complex for this article, we can say that for many purposes, taking the largest six terms, on the three upper left diagonals, provides a sufficiently accurate approximation to the exact product.

Figure 16: Approximating 36 x 487 Using Lattice Multiplication



First Cut: 12,000

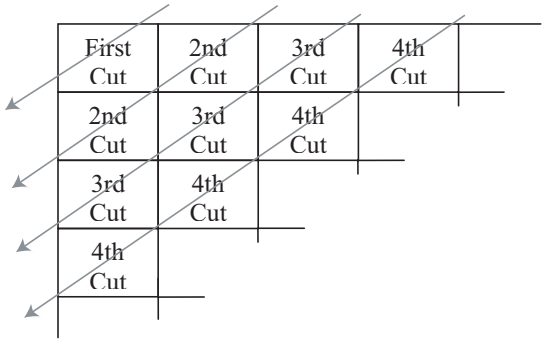
Second Cut: 12,000 + 3,900 = 15,900

Third Cut: 12,000 + 3,900 + 510 = 16,410

The relative error in using 16,410 to approximate the product 16,452 is less than 0.3%.

When more accuracy is needed, a straightforward continuation of the above procedure gives the next terms to add, as illustrated in Figure 17. The upshot is a method to get a (relatively) very accurate approximation to a multidigit multiplication with fairly little work.

Figure 17: Approximating a General Product Using Lattice Multiplication



One final consideration is that because of the rollover phenomenon, discussed in section 2B, we cannot guarantee that the approximate answers computed by the above method will agree in particular decimal places with the exact answer. However, they approximate the exact answer well, in the sense that they produce a small relative error, and in most cases they do have the same largest single-place components as the exact answer.

REFERENCES

[KSF] J. Kilpatrick, J.Swafford and B.Findell, eds. *Adding It Up: Helping Children Learn Mathematics*, National Academy Press, Washington, DC, 2001

[Ma] L.Ma, *Knowing and Teaching Elementary Mathematics*, Lawrence Erlbaum Associates, Mahwah, NJ, 1999.

Appendix: The Rules of Arithmetic

Commutative Rule for Addition: For all real numbers a and b, $a + b = b + a$.

Associative Rule for Addition: For all real numbers a, b, and c, $a + (b + c) = (a + b) + c$.

Commutative Rule for Multiplication: For all real numbers a and b, $ab = ba$.

Associative Rule for Multiplication: For all real numbers a, b, and c, $a(bc) = (ab)c$.

Distributive Rule: For all real numbers a, b, and c, $a(b + c) = ab + ac$.

Existence of Identity for Addition: There is a real number, denoted by 0, with the property that for all real numbers a, $0 + a = a + 0$.

Existence of Identity for Multiplication: There is a real number, denoted by 1, with the property that for all real numbers a, $1 \cdot a = a \cdot 1$.

Existence of Inverse for Addition: Given any real number a, there is a real number, which we denote by $-a$, with the property that $a + (-a) = 0$.

Existence of Inverse for Multiplication: Given any nonzero real number a, there is a real number, which we denote by $\frac{1}{a}$, with the property that $a \cdot \frac{1}{a} = 1$.