Chapter 10

Some Comments on Estimation

Visitors to a museum were told by the curator that a certain artifact was approximately 500,013 years old. When they asked him how the age was determined so precisely, he answered "When I came to work here I was told that it was approximately half a million years old, and I've been working here for 13 years!"

Students in primary school learn to round off numbers to the nearest ten, hundred, and thousand. They are taught the *procedure* of rounding without being told what rounding means, why they would want to round off, or when they should round off. As a result, students have no idea what estimation is about (see the museum curator above). The purpose of this chapter is to give a brief discussion of these issues. The basic lesson here is that mathematics is precise, and it remains so even when it gives out imprecise information, i.e., makes estimates.

The first section gives a precise definition of rounding, and then shows how the usual procedure taught in school follows logically from the definition. It may be observed by now that giving a precise definition of a concept before deriving the procedure associated with the concept is a recurring theme throughout this book. In addition to "rounding", other examples are division-with-remainder, the sum of two fractions, the product of two fractions, the sum of two (positive or negative) rational numbers, etc.

The exposition of this chapter breaks precedent with what we have done so far, which is to be self-contained *in the sense that we explain every concept before we put it to use.* The consideration of estimation requires that we make use of *percent* and *decimals*, concepts we will not take up until Part 2. If we insist on being self-contained, then we would have to postpone this discussion until after Part 2. But estimation, as remarked, makes its appearance in primary school, and a good deal of it makes sense as soon as whole numbers are introduced. For this reason, we take up this topic here, and compromise by asking the reader to skip part of this section if necessary, and revisit it after reading Part 2.

The sections are as follows:

Rounding

Absolute and Relative Errors

Why Make Estimates?

A Short History of the Meter

In the last section, we trace the history of the length of a meter, which is surprisingly relevant to the topic of estimation.

10.1. Rounding

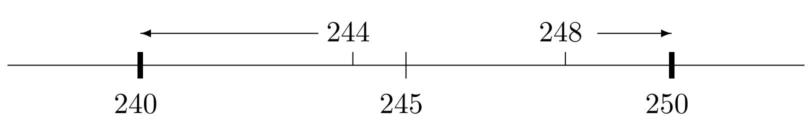
We begin with a precise

Definition. *To* ***round a whole number n to the nearest ten*** *means to replace* n *by the multiple of* 10 *which is closest to* n*. If two multiples of* 10 *are equally close to* n*, the* CONVENTION *is to always choose the bigger number.*

For example, to round 248 to the nearest ten, we look at all multiples of 10 in the neighborhood of 248,

$$\ldots, 210, 220, 230, 240, 250, 260, 270, \ldots,$$

and it is relatively easy to see that 250 is the number we seek. If the number is 244, then the rounded number would be 240. For 245, both 240 and 250 are equally close to 245 and 250 is the bigger of the two, so 245 rounds to 250.



The convention to round to the bigger number in cases like 245 is an American one. Some countries adopt a different convention.

In standard textbooks, students are usually taught to round a number n to the nearest ten by the following algorithm: If the ones digit of n is ≤ 4, change it to 0 and leave the other digits unchanged. But if the ones digit is ≥ 5, then change it to 0 but also increase the tens digit by 1 and leave other digits unchanged. This is correct in most cases, but collapses completely in the case of a number such as 12996. By the definition above, however, it is not difficult to carry out this rounding: since among 12980, 12990, 13000, 13010, ..., the multiple of 10 closest to 12996 is 13000, so the rounded number is 13000.

Nevertheless, there is a correct formulation of the algorithm to round a whole number n to the nearest ten. We can see it from an example such as 12996. We have $12996 = 12990 + 6$, and the rounding to the nearest 10 has nothing to do with 12990 but everything to do with 6: 6 itself rounds to 10, and it follows that 12996 rounds to $12990 + 10 = 13000$, as before. The same reasoning leads to the following general algorithm:

> Write n as $N + \overline{n}$, where $\overline{n}$ is the single-digit number equal to the ones digit of n (and hence N is the whole number obtained from n by replacing its ones digit with 0). Then rounding n to the nearest ten yields the number which is equal to N if $\overline{n} < 5$, and equal to $N + 10$ if $\overline{n} \geq 5$.

Rounding a whole number n to the nearest hundred (resp., thousand, etc.) can be similarly defined. For example, **rounding a whole number *n* to the nearest thousand** means to replace n by the multiple of 1000 which is closest to n. If two multiples of 1000 are equally close to n, the CONVENTION is to always choose the bigger number.

The corresponding algorithm is then:

> Write n as $N + \overline{n}$, where $\overline{n}$ is the three-digit number equal to the last three digits (to the right) of n. Then rounding n to the nearest thousand yields the number which is equal to N if the left digit of $\overline{n}$ is < 5, and equal to $N + 1000$ if the left digit of $\overline{n}$ is ≥ 5.

Activity. Explain to the person sitting nearest you why this algorithm is correct.

We illustrate rounding to the nearest thousand with some examples. Note that the N above is the whole number obtained from n by replacing its last three digits on the right with 000. Note also that the left digit of $\overline{n}$ is the hundreds digit of n. So to round 45,297 to the nearest thousand, we write it (according to the algorithm) as $45000 + 297$. Since the left digit 2 of 297 is < 5, the rounded number has to be 45,000. On the other hand, to round 49,501 to the nearest thousand, we write it as $49000 + 501$. Since the left digit 5 of 501 is ≥ 5, the rounded number is $49000 + 1000 = 50000$. Similarly, rounding 729,998 to the nearest thousand gives 730,000.

Activity. Round 20,245,386 to the nearest ten, nearest hundred, nearest thousand, nearest ten-thousand, and nearest million, respectively. Do the same to 59,399,248.

It remains to point out that the concept of rounding, which has been applied only to whole numbers thus far, is perfectly applicable to decimals as well. We recall a concept from page 99: given a number n (which need not be a whole number), the number kn where k is a *whole number* is called a **multiple** of n, or **whole-number multiple** for emphasis. Now, if a decimal such as 26.8741 is given, then **rounding it to the nearest hundredth (0.01)** means replacing it by the whole-number multiple of 0.01 that is closest to 26.8741.

To find this rounded number, we concentrate on the multiples of 0.01 that are near 26.8741:

$$\ldots, 26.85, 26.86, 26.87, 26.88, 26.89, \ldots.$$

The answer is 26.87, by inspection. More precisely, $26.88 - 26.8741 = 0.0059$, while $26.8741 - 26.87 = 0.0041$. Because $0.0059 > 0.0041$, the rounded number has to be 26.87.

We can also formulate a precise algorithm for this purpose. We need to fix another piece of terminology. Given a decimal such as 26.8741 above, its **first decimal digit** is 8, its **second decimal digit** is 7, its **third decimal digit** is 4, etc. Then

> To round a decimal m to the nearest hundredth, write it as $m = M + \overline{m}$, where M is the decimal whose digits (starting from the left) agree with those of m up to and including the second decimal digit, and are 0 elsewhere. Then the rounded decimal is M if the third decimal digit of $\overline{m}$ is < 5, and it is equal to $M + 0.01$ if the third decimal digit of $\overline{m}$ is ≥ 5.

The decimal $\overline{m}$ is the one obtained from m by replacing every digit of m to the left of its third decimal digit by 0.

Activity. Show that this procedure is correct.

For example, to round 26.8741 to the nearest hundredth, we write it as $26.87 + 0.0041$. Since the third decimal digit 4 of 0.0041 is < 5, the rounded number is 26.87, which confirms the previous answer. As another example, to round 59.99725 to the nearest hundredth, we write it as $59.99 + 0.00725$. Since the third decimal digit 7 of 0.00725 is > 5, the rounded number is $59.99 + 0.01 = 60$.

The formulation of the definition of rounding a decimal to the nearest $0.0\cdots01$, where the 1 is the kth decimal digit, is similar to the case of whole numbers and will be left as an exercise.

Activity. Round 1.70995 to the nearest 10^{-k}, for $k = 1, 2, 3, 4$.

Now we take up the question of why anyone would ever want to round off any number. Consider the population of Berkeley (California) which, according to the city website, is 102,743 (Census 2000). One could raise a legitimate question about how meaningful such a figure is. It is not just the variability of a city's population over time (needless to say, the census was taken sometime before 2000), but even at the moment the census was taken, such a figure had to be inherently inaccurate. Given the unpredictability of death, the mobility of the population (especially considering the large student population), and the acknowledged inability to get all the residents to participate in the census, a conservative estimate is that at least the last three digits, 743, are meaningless. One way the city of Berkeley could have indicated the uncertain nature of this figure would be to avoid any reference to 743 by rounding 102,743 to the nearest thousand. As we have seen, this rounding yields the figure of 103,000. Thus, one can say that the population of Berkeley is about 103,000, and the three zeros at the end of this figure are usually taken to be due to rounding. As a matter of common sense, one would consider "about 103,000" to be a more reasonable description, in context, of the population of Berkeley than "102,743". This is one example of why one wants to round off figures.

One can make a case that even the thousands digit of 102,743 (i.e., 2) is suspect. In that case, we want to eliminate any reference to 2,743 altogether by rounding 102,743 to the nearest ten-thousand. (At the end of the next section, Absolute and Relative Errors, we will explain in a quantitative way why we can afford to round off this way.) Thus to the nearest ten-thousand, the population of Berkeley is about 100,000. Now in everyday conversation, you are more likely to hear "The population of Berkeley is about a hundred thousand" than "The population of Berkeley is about a hundred and three thousand". So the decision to round to the nearest ten-thousand rather than the nearest thousand, while giving a cruder figure, better serves the purpose of "giving a ballpark figure" in everyday life.

As another example of the need for rounding, imagine that you are the leader of a research team and members of your team came up with the result of 58.41672 cm for a particular measurement. But you know that the instrument used in that measurement, while capable of giving readings containing many digits past the decimal point, is accurate only up to the second decimal digit, i.e., only 58.41 is a reliable figure. In this case, you have no choice but to round off to the nearest hundredth to avoid giving out misinformation. In your research report, you therefore give the measurement as 58.42, and the scientific community would understand, as a matter of convention, that the last digit (i.e., 2) involves rounding. This is then another situation where rounding is called for.

10.2. Absolute and Relative Errors

Rounding introduces error. It is therefore imperative that you be aware of the error that comes with the rounding (to the nearest hundred? nearest thousand?) and whether or not the magnitude of the error is something you can live with.

The need of this awareness is well illustrated by the following example. A fourth grade textbook illustrates the usefulness of rounding by looking at an addition problem: $127 + 284 = ?$ After computing directly to get the answer 411, it proceeds to use rounding to check the correctness of 411. It says 127 rounds down to 100 while 284 rounds up to 300, so $127 + 284$ is roughly $100 + 300 = 400$. Since 400 is close to 411, according to this textbook, 411 is a reasonable answer.

The mathematics of the preceding paragraph is extremely flawed and serves as an object lesson on how *not* to teach the usefulness of rounding. To see this, consider the following two questions. First, in saying 411 is close to 400, what is meant by being "close"? Second, should students expect similar closeness each time they estimate the sum of two 3-digit numbers? We begin with the second question.

Suppose we have to add $149 + 147$. Rounding to the nearest hundred gives $100 + 100 = 200$, but the exact answer is of course 296. Right away we see that, no matter how "closeness" is defined, 200 is not close to 296 the same way that 400 is close to 411. Rounding to the nearest hundred is therefore not a good way to check the correctness of additions of 3-digit numbers.

Next, to quantify "closeness", we introduce two standard concepts. The difference between the correct value and an estimated value (always taken to be a positive number) is called the **absolute error**. The ratio

$$\frac{\textit{absolute error}}{\textit{correct value}},$$

which is commonly expressed in percent, is called the **relative error** of the estimation. What the relative error does is to put the absolute error in perspective. Let us explain the last sentence by way of an analogy. Suppose you get two problems wrong on each of two exams, and the two exams have three questions and 15 questions, respectively. Then the number 2 by itself does not tell the full story of your performance; you need to consider that you got $\frac{2}{3}$ of the first exam wrong and but only $\frac{2}{15}$ of the second wrong before you see the difference. This idea of putting the number 2 in the proper context is a basic motivation for introducing the concept of relative error.

For $127+284$, the absolute error is $411-400 = 11$, and the relative error is

$$\frac{11}{411} \approx 3\%,$$

where "$\approx$" means "approximately equal to" and we have rounded off the percent to the nearest one. On the other hand, the estimation of $149 + 147$ as 200 has an absolute error of $296 - 200 = 96$, so that its relative error is

$$\frac{96}{296} \approx 32\%.$$

Although there is no universally accepted definition of the concept of "closeness"—and it must be admitted that what is close or not close depends on the particular situation at hand—one can nevertheless assert that a relative error of more than 10% *generally* would not qualify as "close". The textbook's mistake is not to have forewarned students about the pitfalls of the proposed method. This kind of instruction on the use of estimation is therefore unacceptable.

We now show how one can make use of estimation in a positive way by showing that rounding to the nearest hundred is an effective way to check the addition of 4-digit numbers. Take $4257+3461$. Rounding to the nearest hundred yields an estimate of $4300+3500 = 7800$. Since the correct answer is 7718, the absolute error is $7800 - 7718 = 82$, and the relative error is

$$\frac{82}{7718} \approx 1\%.$$

This shows that the estimation is close to the correct answer, by any definition of "close". We can argue in general. Suppose we have two 4-digit numbers m and n. If we round both numbers to the nearest hundred and add, we introduce an absolute error of at most 100 to the addition, for the following reason. Rounding m alone leads to an absolute error of at most 50 because in choosing among all multiples of 100 for the rounding (see the definition of rounding in the preceding section), m is never more than 50 from a multiple of 100. Same for n. The combined error in the addition is therefore never more than 100. In the case of $4257 + 3461$, we found the absolute error to be 82; one can see easily that for $4250+3450$, the absolute error becomes exactly 100. Now to return to the task at hand, both m and n being 4-digit numbers, the **leading digit** (i.e., the left digit) of m or n is at least 1, so that the leading digit of $m + n$ is at least 2. *Now the leading digit of a 4-digit number is the thousands digit, and therefore $m + n$ is at least 2000.* Thus

$$(m + n) \geq 2000 \quad \text{and therefore} \quad \frac{1}{m+n} \leq \frac{1}{2000}.$$

(We will prove in Chapter 15 that if two whole numbers m, n satisfy $m \geq n > 0$, then $\frac{1}{m} \leq \frac{1}{n}$.) The relative error of the estimation of $m+n$ by rounding m and n to the nearest hundred is then at most

$$\frac{100}{m+n} \leq \frac{100}{2000} = 5\%. \tag{10.1}$$

The conclusion is that

> *if we estimate the sum of two 4-digit numbers by rounding to the nearest hundred, the relative error is at most* **5%**.

As we said, a relative error of 5% gives us confidence that the estimate is reasonable.

In the last argument, we saw the importance of the place value of the leading digit of a number. If the numbers had been 3-digit ones, then all we could have said was that $m + n \geq 200$ and the computation in (10.1) above would have been

$$\frac{100}{m+n} \leq \frac{100}{200} = 50\%.$$

Of course, to have a relative error which might be as great as 50% is nothing to write home about, and this is consistent with the earlier comment that for 3-digit numbers, rounding to the nearest hundred is no way to check the accuracy of addition.

In making estimates, it is critically important to be aware of the relative error of each estimate.

It is time to point out that the place value of the leading digit of a whole number is called the **order of magnitude** of a number. Recall that if the order of magnitude of a whole number is 10^n, then the number has $n+1$ digits. So one may think of order of magnitude as a different indication of "the number of digits" of a number. The order of magnitude of a number is the most basic statement about the size of a number and, not infrequently, the only thing that one cares to know about a number. For example, the national debt as of April 22, 2006, is \$8,379,388,245,684.45, which is of course computed according to a fixed formula and therefore is given down to two decimal places.[1] The order of magnitude is therefore 10^{12}. Eight *trillion* plus. The fact that it is given down to two decimal places is obviously not to be taken seriously: it is a constantly evolving figure that changes with every minute and we as a nation could not care less whether we owe 95 cents more or less. It would be fair to say that "about 8 trillion" is all that matters to an average citizen concerned with this figure.

Let us look at the national debt more closely. How much information do we lose if we only know that it is "about 8 trillion"? Now "8 trillion" implies that *we are rounding to the nearest* 10^{12}. So 8 trillion could be as

[1] It is nevertheless sobering to note that our national debt as of April 23, 2004, was only \$7,141,602,592,641.44.

high as \$8,499,999,999,999 or as low as \$7,500,000,000. The absolute error is at most 5×10^{11}, and the relative error is then at most

$$\frac{5 \times 10^{11}}{7.5 \times 10^{12}} \approx 6.7\%.$$

Thus *by clinging to the simplistic view that the national debt is 8 trillion, which amounts to throwing away 12 digits of a 13-digit number, we are off by only* 6.7%*.* This makes the trade-off of precision for simplicity altogether worthwhile.

Activity. The area of the U.S.A. is 9,629,091 square kilometers. Round it to the nearest thousand, ten-thousand, hundred-thousand, and million. In everyday conversation, which of these rounded figures do you think would be most useful? What is the relative error in that case?

Finally, let us bring closure by returning to our initial example of the population of the city of Berkeley, which is 102,743. The reason that the last four digits 2743 are unreliable has been explained earlier. Let us estimate the relative error of rounding this figure to the nearest 10^4, which then gives a rounded figure of 100,000. Now, although there is no such thing as "the correct value" of Berkeley's population, we can at least make a stab at it. Believing the last four digits of 102,743 to be unreliable, we may assume that the population is, *at the least*, off by 5,000, so that 97,743 is the smallest possible correct population figure for Berkeley. The biggest possible "correct population" of Berkeley is $102743 + 5000 = 107743$, so that the biggest absolute error would be $107743 - 100000 = 7743$. Therefore the relative error of the estimation of the population as 100,000 does not exceed

$$\frac{7743}{97743} \approx 8\%.$$

An error of at most 8% in an everyday context is altogether tolerable.

10.3. Why Make Estimates?

Building on the examples given in the preceding sections, we now give a comprehensive summary of some of the reasons why one would use estimations in place of precise numbers. There are at least three.

(I) *Precision is unattainable or unavailable.* Practically all measurements made in the sciences are approximations. This is because every instrument has built-in limitations to its accuracy. The simplest illustration of this fact is to try to measure the length of your desk by an ordinary metric ruler which has clear markings of millimeters but nothing smaller. Thus the instrument you use for this measurement is only reliable up to one millimeter, i.e., $\frac{1}{10^3}$ of a meter. Therefore, even assuming that your ruler is

100% accurate,[2] your final measurement, given in meters, will be accurate only up to the third decimal digit. If you give a figure with four decimal digits, then the fourth decimal digit could only be an estimate and should be rounded off.

It may be of some interest to note that there is one measurable quantity that is *completely* precise, but that is because of a **fundamental hypothesis** in physics: The speed of light[3] is *exactly* 299,792,458 meters per second, and it is not an estimate because the General Conference on Weights and Measures (CGPM) decided in 1983 to *define* the length of a meter to be 1/299,792,458 of the distance traveled by light in one second. See the last section, A Short History of the Meter on page 150.

Another class of examples of numbers with built-in inaccuracies are astronomical measurements of the distances of stars. You are probably so used to reading about such-and-such a star is x light years away that you do not realize how fantastic it is to be able to make estimates of such astronomical distances. Here we are talking about stars other than the sun, the nearest of which is about 4.3 light years away. Thus if we send a signal to the nearest star other than the sun, it would take 4.3 years to get there. Any thought of a direct measurement is therefore out the window. What estimated distances we come up with have to be the results of extremely sophisticated and indirect inferences. Therefore measurements in an astronomical context are generally estimates and not exact.

Another kind of need for estimations rather than precise values arises from our imprecise use of language. We speak of the "distance" we travel to go to work, for example. We also speak of "the temperature" of a city on a given day. We will leave as an exercise (Exercise 13 at the end of this section) to explain why these are inherently imprecise concepts.

As final examples of the unattainability of precision, we recall the notion of a national debt and the population of a city or state, or for that matter, the nation. These are clearly numbers that have been reluctantly extracted from a sea of ambiguities.

(II) *Precision is unnecessary.* Next time you enter an elevator, look for a sign that says "Capacity x lbs." This x may be 4000, or it may be 2500, but it is hardly likely that if a load of 4001 lbs. is put in the elevator in the former case, the elevator would immediately drop to the bottom of the building. So the main purpose of this number x is more to give a general warning against overloading than a precise determination of carrying capacity. More likely the true capacity has been rounded down from something like 4500 lbs. to 4000 for reasons of safety.

[2]This is of course impossible.

[3]Strictly speaking, it is the speed of light *in a vacuum*.

Consider now the case of weather reporting. Sometimes one hears during freezing winter months announcements about the day's "balmy weather in the sixties". If one insists on precision, such an announcement would be abominable. But precision is hardly called for when the overriding concern is whether you have to go out bundled up in a Siberian overcoat.

Next, we return to the case of the national debt. Consider the following data:

On April 22, 2006, it was \$8,379,388,245,684.45.
On February 23, 1996, it was \$5,017,056,630,040.53.

With these precise numbers in the background, now think about the alternative, *imprecise* rendering of these facts by the simple statement that "our national debt ballooned from 5 trillion to 8 trillion in a span of ten years from 1996 to 2006". Some would say that the latter conveys as much information as a statement giving the precise figures. In a similar vein, it can be argued that for everyday purposes, knowing that the world's population is "about 6 billion people in 2004" is as good as knowing the census figure of 6,610,401,734 people worldwide in April of 2006.

The moral of all these examples is that, sometimes, less is more. Less precision may serve one's needs as well as the full story.

(III) *Estimation is used as an aid to achieve precision.* The most obvious example here is division-with-remainder (Chapter 7), which begins with an estimate—given whole numbers a and b with $b > 0$—of the number q so that $qb \leq a$ but $(q+1)b > a$. Then we obtain the precise value of the remainder of the division as $a - qb$. Note the parallel between estimating the quotient q and rounding a number a to the nearest hundred or nearest thousand (see section 7.4, A Mathematical Explanation (Preliminary), on page 110): both involve trapping a given whole number between two successive multiples of a fixed number, the multiples of b in the case of division-with-remainder and the multiples of 100 or 1000 in the case of rounding to the nearest hundred or nearest thousand. The long division algorithm (Chapter 7) then carries this process of estimation one step further by requiring the estimate of a quotient at each step of the algorithm.

In the last section, we saw how to use rounding *properly* to partially confirm an addition $(4257 + 3461 = 7718)$. Of course estimation can be used to check the other arithmetic operations as well. Let us give a crude example: could $78 \times 86 = 5608$ be correct? No, because $70 \times 80 = 5600$, and 78×86 should be quite a bit bigger than 70×80. Estimations are most powerful for checking order of magnitude. For example, 285×461 cannot possibly be equal to 87385, because rounding 285 down to 250 and 461 to 400 shows that $285 \times 461 > 250 \times 400 = 100{,}000$, so that 285×461 should be at least a 6-digit number.

The need for estimation also arises in other contexts. Suppose you are in a supermarket and you have put seven items in the basket but you only have \$15 with you. It suddenly occurs to you that maybe you do not have enough money. You must make a quick estimate that the seven items do not cost more than \$15. You have neither pen and paper nor calculator with you and you must do the calculations in your head. The price tags read: \$1.25, \$3.25, \$1.39, \$0.99, \$1.49, \$2.42, \$2.79. You round everything to the nearest dollar and add: $1+3+1+1+1+2+3=12$. *But you also know that each rounding introduces an error of \$0.50*, so that if the errors were cumulative, the estimation of the cost could be off by $7 \times 0.5 = 3.5$ dollars. The total cost of the seven items could therefore be as high as $12+3.5=15.5$ dollars. If this were the case, you would be in trouble. This calls for a double-check: the error could be 3.50 only if you have rounded *down* each time, but you actually rounded *up* twice: \$0.99 and \$2.79. Therefore the error is at most $5 \times 0.5 = 2.5$, in which case the total cost is at most $12+2.5=14.5$ dollars. So you are safe. At the cash register, you have your estimate confirmed: the total comes to \$13.58.

The preceding example gives meaning to the following statement made at the beginning of the last section. We use it to conclude this discussion of estimation:

> It is imperative that you be aware of the error that comes with any rounding and whether or not the magnitude of the error is something you can live with.

10.4. A Short History of the Meter

The meter was promulgated to be the unit of length in 1795 under Napoleon's order, and was defined to be 1/10,000,000 (one ten-millionth) of a quarter of the circumference of the earth as measured by the meridian passing through Paris (and the North Pole). This was a grandiloquent decision not grounded in common sense because, given the unevenness of the surface of the earth and the fact that such a meridian is not really a circle (it is more-or-less an ellipse; see Exercise 10 below), this circumference does not lend itself to an easy, precise measurement. Consequently, the length of the meter was, from the very beginning, a fictitious absolute standard at best. In fact, modern measurements yield a value of this quarter circumference about two thousand meters longer than the putative ten million meters. For this reason, the exact value of the meter was revised several times in the past two hundred years.

In the meantime, the question of whether the speed of light in a vacuum is a constant or a variable quantity was heavily debated towards the end of the nineteenth century. The fact that it is constant was first verified to within acceptable limits of accuracy by the Michelson–Morley experiment in

1887, and was adopted as a fundamental hypothesis in physics by Einstein for his theory of special relativity in 1905. When in 1975, the speed of light was measured by lasers to be 299,792,458 meters per second, with an accuracy up to 1.2 meters per second, the radical idea of *recalibrating the meter in terms of the speed of light* began to take hold. The subsequent definition of the meter by the General Conference on Weights and Measures ten years later as 1/299,792,458 of the distance traveled by light in a vacuum in one second was the result.

Chapter 24

On the Teaching of Fractions in Elementary School

The following is a slight revision of an unsuccessful grant proposal written in October of 1999 to NSF-EHR, asking for $50,000 to teach fractions to teachers in the manner of Part 2 of this book.

It is widely recognized that there are at least two major bottlenecks in the mathematics education of grades K–8: the teaching of fractions and the introduction of algebra. Both are in need of an overhaul. I hope to make a contribution to the former problem by devising a new approach to elevate teachers' understanding of fractions.

The need for a better knowledge of fractions among teachers has no better illustration than the following anecdote related by the mathematician Herb Clemens ([**Cle95**]):

> Last August, I began a week of fractions classes at a workshop for elementary teachers with a graph paper explanation of why $\frac{2}{7} \div \frac{1}{9} = 2\frac{4}{7}$. The reaction of my audience astounded me. Several of the teachers present were simply terrified. None of my protestations about this being a preview, none of my "Don't worry" statements had any effect.

This situation cries out for improvement. Through the years, there has been no want of attempts from the mathematics education community to improve the teaching of fractions ([**Lam99**], [**BC89**], [**LB98**], to name just a few), but real success has proven to be elusive. By analyzing these attempts and the existing texts on fractions, for both schools and professional development, one detects certain persistent problematic areas in both the theory and practice, and they can be briefly described as follows:

(1) Most concepts related to fractions are never clearly defined. Fraction, mixed number, decimal, percent, ratio, etc.—each is usually presented by way of a metaphor, e.g., $\frac{1}{3}$ is *like* what you get when you cut a pizza into three "equal" pieces. But mathematics is *precise* and cannot be taught by metaphors alone, or if it is taught by metaphors, then we should not ask students to do precise computations or require them to reason precisely.[1]

(2) The linguistic complexities associated with the common usage of fractions are emphasized from the beginning at the expense of the underlying mathematical simplicity of the concept.

(3) The meaning of each of the four arithmetic operations is almost never given, so that the rules of the operations have to be made up on an ad hoc basis, unrelated to the usual four operations on whole numbers with which students are familiar.

(4) *Mathematical* explanations of essentially all aspects of fractions are lacking.

These four problems are interrelated and are all fundamentally mathematical in nature. For example, if one never gives a clear-cut definition of a fraction, one is forced to "talk around" every possible interpretation of the many guises of fractions in daily life in an effort to overcompensate. A good example is the over-stretching of a common expression such as "a third of a group of fifteen people" into a main theme in the teaching of fractions ([**Moy96**]). Or, instead of offering *mathematical* explanations to children of what the usual algorithms mean and why they are reasonable—a simple task *if* one starts from a precise definition of a fraction—algorithms are justified through "connections among real-world experiences, concrete models and diagrams, oral language, and symbols" ([**Hui98**, page 181]; see also [**LB98**] and [**Sha98**]). It would be so much simpler and so much more to the point if an honest mathematical explanation were given. It is almost as if one makes the concession from the start: "We will offer everything but the real thing".

Let us look more closely at the way fractions are introduced in the classroom. Children are told that a fraction $\frac{c}{d}$, with positive integers c and d, is simultaneously at least five different objects (cf. [**Lam99**] and [**RSLS98**]):

(a) parts of a whole: when an object is equally divided into d parts, then $\frac{c}{d}$ denotes c of those d parts.

[1] See [**Wu10b**].

(b) the size of a portion when an object of size c is divided into d equal portions.
(c) the quotient of the integer c divided by d.
(d) the ratio of c to d.
(e) an operator: an instruction that carries out a process, such as "$\frac{2}{3}$ of".

It is clear that even those children blessed with an overabundance of faith would balk at accepting a concept so magical as to fit all these descriptions all at once. How could this glaring "crisis of confidence" in fractions be consistently overlooked? More importantly, such an introduction to a new topic in mathematics is contrary to every mode of mathematical exposition that is deemed acceptable by modern standards. Yet, even Hans Freudenthal, a good mathematician at the time he switched over to mathematics education, made no mention of this central credibility problem in his Olympian pronouncements on fractions ([**Fre83**]).

Of the existence of such a crisis of confidence, there is no doubt. In 1996, a newsletter for teachers from the mathematics department of the University of Rhode Island devoted five pages of its January issue to "Ratios and Rational Numbers" ([**CRE96**]). The editor wrote:

> This is a collection of reactions and responses to the following note from a newly appointed teacher who wishes to remain anonymous:
>
> "On the first day of my teaching career, I defined a rational number to my eighth grade class as a number that can be expressed as a ratio of integers. A student asked me: What exactly are ratios? How do ratios differ from fractions? I gave some answers that I was not satisfied with. So I consulted some other teachers and texts. The result was confusion ...."

This is followed by three pages worth of input from teachers as well as the editor on this topic, each detailing his or her inconclusive findings after consulting existing texts and dictionaries (!).[2]

In a similar vein, Lamon wrote: "As one moves from whole number into fraction, the variety and complexity of the situation that give meaning to the symbols increases dramatically. Understanding of rational numbers involves the coordination of many different but interconnected ideas and interpretations. There are many different meanings that end up looking alike when they are written in fraction symbol" ([**Lam99**, pages 30–31]). All the while, students are told that no one single idea or interpretation is sufficiently clear to explain the "meaning" of a fraction. *This is a pedagogical disaster,* and to explain why this is so, let us draw an analogy. Suppose you

[2]Allow me to make a trite observation: on technical matters related to science or mathematics, it is usually a fatal mistake to appeal to dictionaries as the ultimate arbiters of truth.

are trying to get driving directions to a small town. Which would you prefer: getting fifty written suggestions on what landmarks to watch out for, what to do at each fork of the road, and how to interpret each road sign along the way, or simply *getting a clearly drawn road map*? So long as nobody believes in giving clear-cut definitions in the teaching of fractions, students will forever "do" fractions without any idea of what they are doing ([**LB98**], [**Lam99**]). After all, what else can they do when they are consistently fed information that is mathematically garbled and incorrect? More pertinent is the following question:

> *Why do we blame students for not learning fractions when it certainly is our fault that we do not teach the subject in a way that can be learned?*

For example, it is certainly difficult for children to be told that a fraction is a piece of pie (according to interpretation (a) above), and then be taught that fractions can be multiplied. *How can you multiply two pieces of pie?*[3]

Sometimes one could "get by" a mathematical concept without a precise definition if its rules of operation are clearly explained. Conjecturally, that was how Europeans in the 16th and 17th centuries dealt with negative numbers. In the case of fractions, however, this is not true even when interpretation (b) of fractions is used. In book after book, fractions are added, multiplied, and divided with no explanation of what the operations mean and with no effort to relate these operations to the same operations on whole numbers. Even when such an attempt is made, the good intention is sabotaged by mathematical flaws, such as the attempt to explain the division of fractions as repeated subtraction (cf. [**PSS00**, page 219]).

That fractions induce math anxiety and mathphobia is no longer news (cf. [**Ash02**]). Informal surveys among teachers also consistently reveal that many of their students simply give up learning fractions at the point of the introduction of addition. It is probably not just a matter of being confused by gcd and lcm, which are central to the traditional way of adding fractions, but more likely a feeling of bewilderment and disgust at being forced to learn a new way of doing addition that seems to bear no relationship to what they already know about addition, namely the addition of *whole numbers*. This then brings us to the problem area (3) at the beginning of this article. We see, for example, that Bezuk and Cramer ([**BC89**, page 156]) willingly concede that

> Children must adopt new rules for fractions that often conflict with well-established ideas about whole numbers.

New rules? In mathematics, one of the ultimate goals is to achieve simplicity. In the context of learning, it is highly desirable, perhaps even mandatory,

[3] I have appropriated this question from a delightful article by Kathleen Hart in [**Har00**].

that we convey this message of simplicity to students. However, when we tell students that a concept as simple as *addition* must be different for whole numbers and fractions, we are certainly teaching them incorrect mathematics (see the introduction of Chapter 14). Even when students are willing to suspend disbelief to go along on such a weird journey, they pay a dear price. Indeed, there are recurrent reports of students at the University of California at Berkeley and at Stanford University claiming in their homework and exam papers that $\frac{a}{b} + \frac{a}{c} = \frac{a}{b+c}$ and $\frac{a}{b} + \frac{c}{d} = \frac{a+c}{b+d}$.

All in all, a mathematician approaching the subject of fractions in school mathematics cannot help but be struck by the total absence of the characteristic features of mathematics: precise definitions as starting point, logical progression from topic to topic, and most importantly, explanations that accompany each step. This is not to say that the teaching of fractions in elementary school should be rigidly formal from the beginning. Fractions should be informally introduced no later than the second grade (because even second graders need to worry about drinking "half a glass" of orange juice!), and there is no harm done in allowing children to get acquainted with fractions in an intuitive manner up to about the fourth grade. An analogy may be helpful here. The initial exploration of fractions may be taken as the "data-collecting phase" of a working scientist: just take it all in and worry about the "how and why" later. In time, however, the point will be reached when said scientist must sit down to organize and theorize about his or her data. So it is that when students reach the fifth grade ([**CAF99**]) or the sixth grade ([**PSS00**]), their mathematical development cannot go forward unless "miracles", such as having one object $\frac{c}{d}$ enjoying the five different properties of (a)–(e) above, are fully explained and rules such as $\frac{a}{b}/\frac{c}{d} = \frac{ad}{bc}$ are justified. And it is at this critical juncture of students' mathematical education that I hope to make a contribution.

The work done on the teaching of fractions thus far has come mainly from the education community. Perhaps because of the recent emphasis on situated learning, discussions of the teaching of fractions in the education research literature tend to stay *at the source*, in the sense that attention is invariably focused on the interpretation of fractions in a "real world" setting. Such an emphasis seems to ignore the fact that when fractions are presented to beginners in a variety of contexts with the attendant myriad interpretations, they tend to get confused. Because students are not given the mathematical underpinnings of these interpretations, they end up being exposed only to the raw data but not the theory-building that makes sense of fractions (see [**Wu08**]). They are denied access to learning about an essential component of doing mathematics, namely, *when confronted with complications, try to abstract in order to achieve understanding.* Even children should be exposed to correct mathematical thinking as early as possible, and in a manner as simple as possible. Students' first serious encounters with the computation of fractions—generally in the fifth and sixth grades—would be the right moment in the school curriculum to begin emphasizing the abstract component of mathematics and make the abstraction a key point of classroom instruction (again, see [**Wu08**]). By so doing, one would also be giving students a head start in their quest to learn algebra. The ability to abstract, so essential in algebra, should be taught gently and early in the school curriculum, which would mean during the teaching of fractions. By giving abstraction its due in teaching fractions, we would be easing students' passage to algebra as well (cf. [**Wu01**], [**NMP08a**], [**NMP08b**], [**Wu09a**]).

It takes no insight to conclude that two things have to happen if mathematics education in K–8 is to improve: there must be textbooks that treat fractions logically, and teachers must have the requisite mathematical knowledge to guide their students through this rather sophisticated subject. I propose to take up the former problem by writing a book to improve teachers' understanding of fractions.

The first and main objective of this book is to give a treatment of fractions and decimals for teachers of grades 5–8 which is mathematically correct in the sense that everything is explained and the explanations are sufficiently elementary to be understood by elementary school teachers. In view of what has already been said above, an analogy may further explain what this book hopes to accomplish. Imagine that we are mounting an exhibit of Rembrandt's paintings, and a vigorous discussion is taking place about the proper lighting to use and the kind of frames that would show off the paintings to best advantage. Good ideas are also being offered on the printing of a handsome catalogue for the exhibit and the proper way to publicize the exhibit in order to attract a wider audience. Then someone takes a closer look at the paintings and realizes that all these good ideas might come to naught because some of the paintings are fakes. So finally people see the need to focus on the most basic part of the exhibit—the paintings—before allowing the exhibit to go public. In like manner, what this book would try to do is to call attention to the need of putting the *mathematics* of fractions in proper order before considering pedagogical strategies and classroom activities in the actual teaching. See [**Wu08**] for further discussions along this line.

Looking ahead, this book—and similar volumes for middle school and high school teachers—which tries to pull the teaching of fractions and rational numbers into the mathematical mainstream, can only be a beginning. Much remains to be done in terms of bringing mathematical integrity back to school textbooks, pre-service professional development, education research, and the school mathematics curriculum. All this requires long-term collaborations between mathematicians and educators. A general discussion of the framework for such a collaboration is given in [**Wu06**].